Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 37, NUMBER 11 NOVEMBER, 1947

The Ultimate Sensitivity of Radiation Detectors

R. CLARK JONES
Research Laboratory, Polaroid Corporation, Cambridge, Massachusetts
(Received June 23, 1947)

An upper limit is obtained for the sensitivity of any radiation detector which satisfies Lambert's law and whose effective area is independent of the wave-length of the radiation. Expressions for the minimum detectable power are given by Eqs. (3.8) and (5.4).

The treatment given is based on thermodynamics, and makes no use of statistical mechanics. In the special case in which the fluctuations in the output of the detector are due entirely to fluctuations in the radiation, the results obtained agree with those obtained from a statistical mechanical treatment of the fluctuations in the radiation.

The treatment of the first five sections is confined to detectors, called "temperature detectors", which operate by means of the change in their temperature brought about by the radiation. Such detectors include the thermocouple, the bolometer, and the radiometer. In Sections VI and VIII it is shown that the results hold for any detector satisfying the conditions in the first paragraph, provided that the emissivities are interpreted as quantum efficiencies. Section VII indicates that if the condition regarding Lambert's law is given up, then there exists no lower limit on the minimum detectable power. Section IX is devoted to numerical examples and a comparison with the reported sensitivities of bolometers.

INTRODUCTION

THE existence of statistical fluctuations sets an upper limit, which depends on the temperature, on the sensitivity of any device which detects electromagnetic radiation, including radiant heat.

The statistical fluctuations discussed here are those caused by temperature fluctuations of the detector and/or caused by statistical fluctuations in the radiation field. In actual radiation detectors, however, the voltage fluctuations which constitute Johnson noise frequently determine the limiting sensitivity. In these cases Johnson noise is the limiting factor only because the coupling between the radiation-detecting element and the electrical circuit is too weak. In the case of the bolometer, for example, the signal-to-Johnson noise ratio is proportional to the temperature coefficient of the resistance of the bolometer element. Because there is no fundamental limitation on possible values of the temperature coefficient of resistance, it follows that there is no fundamental limit on the signal-to-Johnson noise ratio. Similarly, in thermocouple detectors there is no fundamental limitation on the Peltier coefficient, so that again the signal-to-Johnson noise ratio has no upper limit. In both of these cases, however, it will be shown in this paper that the fluctuations in the temperature of the heat-detecting element serve to provide a fundamental limit on the signal-to-noise ratio, which is not greatly in excess of the signal-to-Johnson noise ratios which are now obtainable.

Both the bolometer and the thermocouple are examples of "temperature detectors", defined here as detectors which operate through the mechanism of the change in their temperature produced by the radiation to be detected. Such

detectors are also suitable for use as temperature-measuring devices. An example of a radiation detector which is not a temperature detector is the photoelectric cell, or the dipole antenna.

The results obtained in this paper hold for all types of radiation detectors whose size is large compared with the wave-lengths of the radiations being detected, and which obey Lambert's law. In the case of temperature detectors, the results apply to those which are conductively cooled, as well as to bolometer elements which are cooled primarily by radiation to their surroundings.

The theoretical value of the minimum detectable power will be computed first on the basis of the assumption that the detector is a temperature detector. In the case of such a temperature detector which has no conductive contact with its surroundings, and thus communicates with its environment only by electromagnetic radiation, the fluctuation in its temperature would be zero except for the fluctuation in the radiation itself. Accordingly, it may be concluded that the fluctuations in the output of such a detector are due solely to fluctuations in the radiation, and will accordingly be the same for any detector, whether or not the detector operates through measuring the change in its temperature.

Thus the results obtained in the paper hold not only for bolometers and thermocouples, but also for photoelectric cells, photo-conductive and photolytic cells, and any other detector whose dimensions are large compared with the wave-lengths being detected, and which obey Lambert's law.

In order to provide a clear distinction between Johnson noise and the noise at the output of the detector which is due to fluctuations in the radiation or to temperature fluctuations, the latter type of noise will be called temperature noise* throughout this paper.

Two different frequencies, denoted by ν and f, are used. The frequency ν is the frequency of the electromagnetic radiation itself, and f is the frequency of the modulation of the radiant power incident upon the detector.

* The term thermal noise would be preferred, but for the fact that this term is already in wide use as a synonym for Johnson noise.

I. ELEMENTARY BOLOMETER THEORY

This section contains no original material and serves only to introduce the notation and the framework of ideas which are used in the following sections.

Consider a material body of thermal capacity C which is in thermal equilibrium with a much larger body whose temperature is T_0. It is assumed that the dynamic behavior of the body may be specified in terms of a single time constant, so that the cooling constant κ of the material body with respect to its surroundings may be defined by the following differential equation for the approach to equilibrium with a large body, when the temperature of the small body is temporarily displaced for equilibrium:

$$C(d\theta/dt)+\kappa\theta=0. \tag{1.1}$$

In this expression t is the time and θ is the difference between the instantaneous absolute temperature T of the body and its equilibrium temperature T_0:

$$\theta\equiv T-T_0. \tag{1.2}$$

Suppose now that because of a change in the environment of the body the flow of heat to the body (when its temperature is T_0) is increased by the amount H, where H may be a function of the time. The differential Eq. (1.1) then becomes

$$C(d\theta/dt)+\kappa\theta=H(t). \tag{1.3}$$

Now suppose that the flow of heat H oscillates sinusoidally about the value zero with a frequency f:

$$H=H_0\cos 2\pi ft. \tag{1.4}$$

By Eq. (1.3), the amplitude of the steady-state oscillation of the temperature is then given by

$$\theta_0=\frac{H_0/\kappa}{[1+(2\pi f\tau)^2]^{\frac{1}{2}}}, \tag{1.5}$$

where the time constant τ is defined by

$$\tau\equiv C/\kappa. \tag{1.6}$$

In terms of τ the differential Eq. (1.3) may be rewritten in the following form:

$$\tau(d\theta/dt)+\theta=(H(t)/\kappa). \tag{1.7}$$

II. TEMPERATURE FLUCTUATIONS

A thermodynamic treatment will now be used to obtain an expression for the mean-square fluctuation of the temperature of the material body with thermal capacity, C, about the temperature T_0 of the large body with which it is in thermal equilibrium.

Let l be any property of a body, and let $\bar{l}$ be its mean value at thermodynamic equilibrium. Then it may be shown[1] that

$$\langle (l-\bar{l})^2 \rangle_{Av} = -k \Big/ \frac{\partial^2 S}{\partial l^2}, \tag{2.1}$$

where S is the entropy of the body and where k is Boltzmann's constant. Now let l be the temperature difference θ. Then when H is zero one finds from Eq. (2.1)

$$\theta_{\text{r.m.s.}}{}^2 = \langle \theta^2 \rangle_{Av} = -k \Big/ \frac{\partial^2 S}{\partial T^2}. \tag{2.2}$$

One has

$$\frac{\partial S}{\partial T} = \left(\frac{1}{T}\right)\left(\frac{\partial Q}{\partial T}\right) = \frac{C}{T}. \tag{2.3}$$

Upon differentiating (2.3) with respect to the temperature and substituting the result in (2.2), one obtains, finally,

$$\theta_{\text{r.m.s.}}{}^2 = kT^2/C. \tag{2.4}$$

It is perhaps worth while to pause in order to emphasize the generality of the result expressed in Eq. (2.4). This relation follows from the assumption that the body with thermal capacity C is in thermal equilibrium with a much larger body but makes no mention of the tightness of the coupling between the two bodies. If, for example, the two bodies were separated by a layer with poor conductivity, the frequency spectrum of the fluctuations would be quite different from the spectrum in the case of good conductivity. By itself, however, thermodynamics has nothing to say about the nature of the frequency spectrum of the fluctuations, and in order to obtain the frequency spectrum of the fluctuations further analysis is necessary.

Equation (2.4) may also be written in the convenient form

$$\theta_{\text{r.m.s.}}{}^2 = kT^2/(\kappa\tau). \tag{2.5}$$

[1] Paul S. Epstein, *Textbook of Thermodynamics* (John Wiley and Sons, Inc., New York, 1937), pp. 389–390.

III. MINIMUM DETECTABLE POWER OF A RADIATION DETECTOR

In this section the results derived in Section II are used to derive an expression for the minimum detectable power of any radiation detector which operates by virtue of the change in its temperature which is induced by the signal which it is to detect.

Let the body whose thermal capacity is C be the detector. Furthermore, let H_m be the minimum detectable power of the detector, defined as the steady power *incident* upon the entire area of the detector which causes a change in its temperature equal to $\theta_{\text{r.m.s.}}$. This definition serves to introduce precision to a concept which is otherwise somewhat vague. In any event, however, the minimum detectable power just defined will be close to the value obtained by any reasonable criterion. Provided that the fluctuations in temperature are converted into changes in voltage by a proportionality factor which is independent of frequency, the definition of H_m just given means that the signal voltage will be equal to the r.m.s. voltage of the temperature noise. The case in which this condition is not satisfied is considered below in Section V.

Since the power *absorbed* by the detector is $\epsilon_s H_m$, where ϵ_s is the effective emissivity (absorption coefficient) of the surfaces of the detector for the signal in question, one finds from Eq. (1.3) that

$$H_m = \kappa\theta_{\text{r.m.s.}}/\epsilon_s. \tag{3.1}$$

By the use of (2.5), Eq. (3.1) may be written

$$H_m{}^2 = \kappa kT^2/(\epsilon_s{}^2\tau). \tag{3.2}$$

In order to make the minimum detectable power as small as possible at a given temperature, it follows from the last equation that it is desirable to make the time constant τ as long as possible, to make ϵ_s equal to unity, and to make the cooling constant κ as small as possible.

Equation (3.2) is the general expression for the sensitivity of a detector which operates by the mechanism of a change in its temperature. Consideration will now be given to the value of the cooling constant κ.

If H is the amount of heat absorbed per unit time by the detector, the cooling constant κ may be defined by

$$\kappa = dH/dT, \tag{3.3}$$

where in taking the derivative it is understood that the change in the temperature T of the detector takes place so slowly that the derivative has its maximum value.

The thermal radiation of a detector sets a lower limit to the value of the cooling constant κ. Provided that the detector satisfies Lambert's law, so that it is possible to define an emissivity which is independent of the angle of incidence of the radiation upon the detector, the quantity H defined in the last paragraph may be written

$$H=A\epsilon_r\sigma T^4, \tag{3.4}$$

where A is the total area** of the detector, where σ is the Stefan-Boltzmann radiation constant, and where the quantity ϵ_r is the effective emissivity of the detector for thermal radiation of temperature T. Equation (3.4) thus yields the following explicit expression for the cooling constant κ:

$$\kappa=A\sigma\frac{d(\epsilon_r T^4)}{dT}, \tag{3.5}$$

where it has been assumed that the area A is independent of the temperature.

It is convenient to define a third effective emissivity ϵ_u, defined by the relation

$$\kappa=4A\epsilon_u\sigma T^3, \tag{3.6}$$

so that ϵ_u may be written

$$\epsilon_u=\frac{1}{4T^3}\frac{d(\epsilon_r T^4)}{dT}=\frac{d(\epsilon_r T^4)}{dT^4}. \tag{3.7}$$

The substitution of (3.6) in (3.2) then yields the following expression for the minimum detectable power of a detector whose cooling is purely radiative:

$$H_m^2=4A\cdot\sigma T^4\cdot kT\cdot\epsilon_u/(\epsilon_s^2\tau). \tag{3.8}$$

It follows from this relation that in order to make the minimum detectable power at a given temperature as small as possible, the area of the bolometer should be made small, the time constant should be made long, and the ratio ϵ_u/ϵ_s^2 should be made as small as possible.

** If the bolometer is everywhere convex, then the area A is the geometrical area of the surface of the bolometer. If, however, the surface is such that not every point to it can be touched by a tangent plane, then the area A is the minimum area of an envelope enclosing the bolometer such that every point of the envelope can be touched by a tangent plane.

IV. EVALUATION OF THE EFFECTIVE EMISSIVITIES

The effective emissivity ϵ_s can be expressed in terms of the power spectrum of the signal and in terms of the frequency dependence of the emissivity. Let $H_s(\nu, r)$ be the power per unit frequency band width per unit area incident upon the detector at the point r at the frequency ν. The frequency ν is the frequency of the electromagnetic radiation. Similarly, let $\epsilon(\nu, r)$ be the emissivity of the detector at the point r at the frequency ν. Then the effective emissivity ϵ_s may be defined

$$\epsilon_s=\frac{\int dA\int H_s(\nu, r)\epsilon(\nu, r)d\nu}{\int dA\int H_s(\nu, r)d\nu}=\frac{\int dA\int H_s(\nu, r)\epsilon(\nu, r)d\nu}{H_m}. \tag{4.1}$$

In these expressions the integral with respect to A should be extended over the total area of the detector, and the limits on the integral with respect to the frequency ν are zero and infinity. Evidently, if $\epsilon(\nu, r)$ is independent of ν and r, then the effective emissivity ϵ_s is equal to ϵ.

The effective emissivity ϵ_r may be defined in terms of the power spectrum of the thermal radiation and the frequency dependence of the emissivity of the detector in exactly the same way employed in deriving Eq. (4.1). Let $J(\nu, T)$ be the power per unit frequency band of blackbody radiation of temperature T and frequency ν incident upon unit area of a surface, and let $\epsilon(\nu, r)$ be the same quantity previously defined. Then the effective emissivity ϵ_r may be defined by

$$\epsilon_r=\frac{\int dA\int J(\nu, T)\epsilon(\nu, r)d\nu}{\int dA\int J(\nu, T)d\nu}=\frac{\int dA\int J(\nu, T)\epsilon(\nu, r)d\nu}{A\sigma T^4}, \tag{4.2}$$

where, just as in the case of Eq. (4.1), the integral with respect to A is to be extended over the surface of the detector, and the integral with respect to frequency should extend from zero to infinity.

The substitution of (4.2) in (3.5) now yields the following expression for the cooling constant κ:

$$\kappa = \int dA \int \frac{\partial J(\nu, T)}{\partial T} \epsilon(\nu, r) d\nu, \qquad (4.3)$$

where it has been assumed that $\epsilon(\nu, r)$ is independent of the temperature of the detector.

By Eq. (3.6) the effective emissivity ϵ_u is now equal to

$$\epsilon_u = \frac{\int dA \int \frac{\partial J(\nu, T)}{\partial T} \epsilon(\nu, r) d\nu}{4A\sigma T^3}. \qquad (4.4)$$

The theory of thermal radiation, as developed by Planck, yields the following expression[2] for the function $J(\nu, T)$:

$$J(\nu, T) = \frac{2\pi h \nu^3}{c^2 (e^{h\nu/kT} - 1)}, \qquad (4.5)$$

where h is Planck's constant and c is the velocity of light in a vacuum. The derivative of J with respect to T is then

$$\frac{\partial J(\nu, T)}{\partial T} = \frac{2\pi h^2 \nu^4}{c^2 k T^2} \frac{e^{h\nu/kT}}{(e^{h\nu/kT} - 1)^2}. \qquad (4.6)$$

At this point it is convenient to introduce a dimensionless frequency defined by

$$\xi \equiv h\nu/kT. \qquad (4.7)$$

If, furthermore, $\epsilon(\nu, r)$ is expressed in terms of this new variable so that it becomes $\epsilon(\xi, r)$, the cooling constant κ as defined by (4.3) may now be written

$$\kappa = \frac{2\pi k}{c^2} \left(\frac{kT}{h}\right)^3 \int dA \int_0^\infty \frac{\xi^4 e^\xi}{(e^\xi - 1)^2} \epsilon(\xi, r) d\xi. \qquad (4.8)$$

Accordingly, the effective emissivity ϵ_u may be written

$$\epsilon_u = \frac{15}{4\pi^4 A} \int dA \int_0^\infty \frac{\xi^4 e^\xi}{(e^\xi - 1)^2} \epsilon(\xi, r) d\xi, \qquad (4.9)$$

with the help of the following relation between σ and the other constants:

$$\sigma \equiv \frac{2\pi^5}{15} \frac{k^4}{c^2 h^3}. \qquad (4.10)$$

The effective emissivities, ϵ_s and ϵ_u, which enter into the relation (3.7) have now been evaluated as explicitly as possible in the absence of more detailed information about the power spectrum of the signal $H_s(\nu, r)$ and the frequency dependence of the emissivity, $\epsilon(\nu, r)$. The results are expressed in Eqs. (4.1) and (4.9). In the special case in which $\epsilon(\nu, r)$ is independent of both the frequency ν and the position r, and is accordingly equal to a constant ϵ_0, both ϵ_s and ϵ_u become equal to ϵ_0. In the case of Eq. (4.1), this result follows by definition, but the result is not so immediately evident from Eq. (4.9). In this case, the equivalence of ϵ_u and ϵ_0 depends upon the following integral relation:

$$\int_0^\infty \frac{\xi^4 e^\xi}{(e^\xi - 1)^2} d\xi = 24\zeta(4) = \frac{4\pi^4}{15}, \qquad (4.11)$$

where $\zeta(z)$ is the Riemann zeta-function.[3]

V. ARBITRARY FREQUENCY RESPONSE

The relations developed so far have involved the assumption that the thermal characteristics of the detector may be expressed in terms of a single time constant; furthermore, in order that the definition of the minimum detectable power be a realistic one, it is necessary to suppose that the proportionality factor between the electrical output of the detector and the change in its temperature is independent of frequency. The possession of a single time constant, however, is usually the case only with detectors whose cooling is purely radiative. In the case of detectors which are cooled primarily by conduction as, for example, with the high speed bolometers developed during the war, the thermal behavior of the detecting element cannot be expressed in terms of a single time constant. Accordingly, it is desirable to generalize the expression (3.7) so that it holds for a detecting element which, in combination with its associated electrical amplifiers, has any specified frequency response.

[2] J. K. Roberts, *Heat and Thermodynamics* (Blackie and Son, Ltd., London, 1933), second edition, Chapter 19.

[3] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, Teddington, England, 1935), fourth edition, Chapter 13.

The fact indicated by Eq. (2.5), that the mean-square temperature fluctuation is inversely proportional to the time constant τ, means that the power spectrum of the thermal fluctuations is independent of frequency, except as modified by the frequency response of the detector. Accordingly, if the frequency response of the detector plus its associated amplifiers is expressed by $Z(f)$ with the normalizing condition that the value of $Z(f)$ at zero frequency be unity, then the expression $1/\tau$, which occurs in (2.5) and in (3.7), should be replaced by

$$1/\tau \rightarrow 4\int_0^\infty |Z(f)|^2 df. \tag{5.1}$$

In the case of a detector with a single time constant, the function $Z(f)$ has the form

$$Z(f)=\frac{1}{1+2\pi i f\tau}, \tag{5.2}$$

and in this case the expression on the right in (5.1) is equal to $1/\tau$.

The significance of the cooling constant κ remains unchanged; it is the constant effective in the limit of very low frequencies.

If the minimum detectable power of a sinusoidally oscillating signal, whose power varies sinusoidally, is defined by the condition that half of the peak-to-peak signal voltage is equal to the r.m.s. temperature-noise voltage, then the minimum detectable power of a sinusoidal signal with frequency f, as measured by half of its peak-to-peak amplitude, is given by

$$H_m^2=\frac{4\kappa kT^2}{\epsilon_s^2}\frac{\int_0^\infty |Z(f)|^2 df}{|Z(f)|^2}. \tag{5.3}$$

When κ has the minimum value given by Eq. (3.6), the last relation becomes

$$H_m^2=16A\cdot\sigma T^4\cdot kT\cdot\frac{\epsilon_u}{\epsilon_s^2}\frac{\int_0^\infty |Z(f)|^2 df}{|Z(f)|^2}. \tag{5.4}$$

In the last two equations the expressions are independent of the normalization of the function $Z(f)$, so that the equations are in a form suitable for use with detector-amplifier systems with zero d.c. response.

The expressions (5.3) and (5.4) represent the minimum detectable power, as measured by half of its peak-to-peak amplitude, defined by the condition that the r.m.s. noise voltage is equal to the signal voltage, as measured by half of its peak-to-peak value. The same expressions are obtained, no matter what measure of the incident power and of the signal voltage are used, provided only that they be the same. On the other hand, however, it is essential that the noise voltage be measured by its r.m.s. value. Other measures of the noise voltage may be used only if corresponding changes in (5.3) and (5.4) are made.

It should perhaps be emphasized again that it is explicitly assumed that the level of Johnson noise is so much lower than that of the temperature noise that the effect of Johnson noise in raising the minimum detectable power may be ignored. In the event that the level of the Johnson noise may not be ignored, the r.m.s. Johnson-noise voltage should be computed independently, and the total r.m.s. noise voltage obtained by taking the square root of the sum of the mean-square Johnson-noise voltage and the mean-square temperature-noise voltage.

VI. THE RESULTS HOLD FOR ANY RADIATION DETECTOR

The general expressions (3.8) and (5.3) for the minimum detectable power of a radiation detector have been derived explicitly for a detector which operates through the mechanism of the change in its temperature brought about by the signal to be detected.

In the case of the detector which is cooled only by radiation, however, the fluctuations of temperature would necessarily be zero if there were no fluctuations in the transfer of heat by the radiation. Accordingly, it may be concluded that the fluctuations in the output of the detector are due to fluctuations in the radiation, and would, therefore, be the same for any radiation detector, no matter what its mechanism might be. This statement is qualified only by the assumptions which have been made in the derivation, the most important ones being the assumption that the effective area is independent of the

frequency of the radiation and of the temperature, and that the absorption coefficient of the detector is independent of the angle of incidence.

The conclusions of the last paragraph are supported by the fact that the writer has been able to derive the expressions (3.8) and (5.4) by an altogether different method in which the fluctuations are attributed to statistical fluctuations in the absorption and emission of the light quanta of the radiation field. This derivation employs the statistical mechanics of Bose-Einstein assemblies and theory of power spectra employed by Rice[4] for the treatment of Johnson noise. This derivation will not be given here, because it yields no information not provided by the present treatment, and because it is less general in that it provides no information about the fluctuations of the detector which is cooled by other means than radiation.***

Since the fluctuations in a radiation-cooled detector are due entirely to the fluctuations in the incident radiation, it follows from the material presented that the mean-square fluctuation per unit frequency band width in the power incident upon a detector of area A is given by

$$16A \cdot \sigma T^4 \cdot \kappa T. \tag{6.1}$$

Presumably this relation holds for all frequencies small compared with the reciprocal of the period required for the absorption or emission of the quanta.

VII. A FURTHER REDUCTION IN THE MINIMUM DETECTABLE POWER

It has been assumed up to this point that the quantum efficiency of the detector is independent of the direction of incidence of the radiation upon the detector. If this condition is given up, then it no longer is possible to provide an expression for the minimum detectable power which does not depend upon the degree of departure from Lambert's law. The signal-to-noise ratio of an array of antennas, for example, is proportional to the gain of the array, when the signal is incident from the direction of maximum response.

Even in the case of temperature detectors whose blackened surfaces provide a fairly close approximation to Lambert's law, it is possible by the use of radiation reflectors to produce a detector which exhibits a marked departure from Lambert's law, and thus permits one to obtain a lower minimum detectable power than that predicted by Eq. (5.4). A housing should be placed around the detector whose inner surface is as nearly as possible a perfect reflector for the frequencies to which the detector is sensitive; the housing must be incomplete in those directions from which radiation is to be detected. This housing will evidently reduce the cooling constant of the detector, because a fraction of the radiation which leaves the detector will return to the detector.

From the point of view used in the treatment by statistical mechanics, the reduction in the minimum detectable signal occurs because the photons which leave the detector and are reflected back by the reflector are absorbed and emitted at times so closely adjacent that the emitted and absorbed photons make no appreciable contribution to the low frequency components of the power spectrum of the radiation.

Evidently the degree by which the minimum detectable power may be reduced by the use of radiation reflectors will depend on the details of the geometry of the system and the degree to which the reflector approaches unity reflectance. These matters will not be discussed here.

The use of radiation shields yields necessarily a detector which does not obey Lambert's law.

VIII. SIGNIFICANCE OF THE EMISSIVITIES AS QUANTUM EFFICIENCIES

The reader may at first be disturbed by the fact that the general expression (3.8) for the minimum detectable power depends not only on the ratio, but also on the absolute values, of the emissivities. This situation occurs because the contribution to the signal voltage of the light quanta which constitute the signal is proportional to their number, whereas the temperature-

[4] S. O. Rice, Bell. Syst. Tech. J. **23**, 282 (1944).

*** Just before the manuscript of this paper was sent to the editor, the writer's attention was called to a paper by W. B. Lewis, "Fluctuations in streams of thermal radiation," Proc. Phys. Soc. London **59**, 34 (1947). This paper contains essentially the treatment by statistical mechanics described above. Actually, the treatment in Lewis' paper does not consider the question of emissivities, and accordingly is restricted to detectors whose quantum efficiency is independent of the frequency of the radiation. This restriction is not fundamental, since the emissivities may very easily be inserted in Lewis' treatment, with the same result at that obtained here.

noise voltage is proportional to the square root (aside from slight corrections because of the Bose statistics of the quanta) of the number of the quanta in the random radiation field. Thus if the number of quanta reaching and leaving the detector were all reduced by a constant factor, then the minimum detectable power would be increased.

Therefore, the significance of the emissivities for an arbitrary detector is that of the quantum efficiency of that detector. In the case of the temperature detector, the emissivity is simply the quantum efficiency for the absorption of incident light quanta. In the case of the photoelectric cell, for example, the quantum efficiency is the ratio of the number of photoelectrons produced and utilized to the number of incident quanta of the frequency in question.

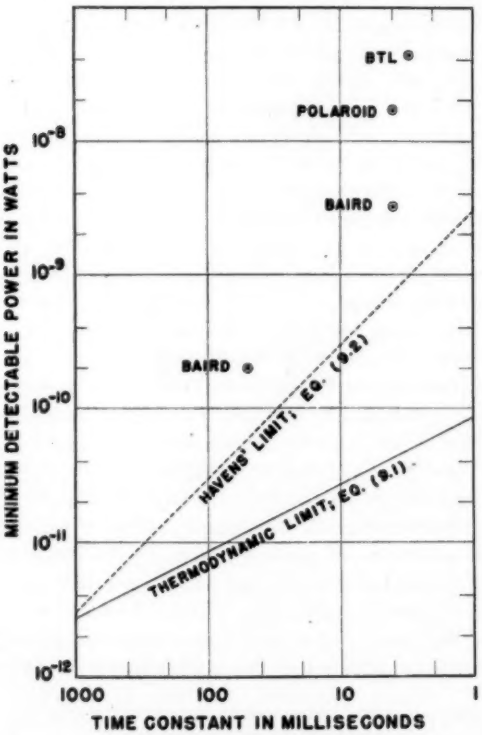

FIG. 1. The solid line shows the theoretical, minimum detectable power for a derector whose quantum efficiency is unity and whose effective area is one square millimeter at all wave-lengths, as predicted by Eqs. (3.8) and (5.4). The detector is assumed to have a temperature of 300°K and to obey Lambert's law. The dashed line shows the minimum detectable power, according to an engineering estimate made by R. J. Havens of what might be done in constructing bolometers with current materials and techniques. The four circles show the minimum detectable powers, computed on the basis specified in the Appendix from performance data supplied by their designers, of four bolometers constructed during or since the last war.

IX. NUMERICAL EXAMPLES

The Heat Detector

Consider first a temperature detector, such as a bolometer, whose emissivity is unity for all frequencies. Suppose that its total surface area is one square millimeter, that its temperature is room temperature (300°K), and that it is cooled entirely by radiation. Then if it is not equipped with radiation shields, but is exposed to the radiation field in every direction, so that it obeys Lambert's law, Eq. (3.8) yields the following expression for the minimum detectable power:

$$H_m = \frac{2.76\times10^{-12}\ \text{watt}}{\tau^{\frac{1}{2}}}, \qquad (9.1)$$

where τ is the time constant in seconds, and where the following numerical values are used:

$$\sigma T^4 = 4.64\times10^{-2}\ \text{watt/cm}^2,$$
$$kT = 4.11\times10^{-21}\ \text{joule},$$
$$4A = 0.04\ \text{cm}^2.$$

Equation (9.1) is shown graphically in Fig. 1. For purposes of comparison, the sensitivity of a few modern bolometers is also shown on the plot. Because, however, of the present confused state with regard to the sensitivities of heat detectors, the writer has felt it necessary, in order to avoid adding to the confusion, to include in the Appendix a statement of exactly how the stated sensitivities were computed.

In examining Fig. 1 it must be remembered that the sensitivity of existing bolometers and thermocouples is limited not by the radiation fluctuations alone involved in Eq. (9.1), but rather by Johnson noise.

The dashed line, represented by the formula

$$H_m = \frac{3.0\times10^{-12}\ \text{watt}}{\tau}, \qquad (9.2)$$

(where τ is in seconds) is the minimum detectable power of a bolometer operated at room temperature, as limited by Johnson noise, according to

an estimate made by Havens[5] on the basis of current techniques and materials now available. This curve also refers to an area of one square millimeter. Havens' limit is by no means intended to be a fundamental limit which cannot be exceeded, but has rather the nature of an optimistic engineering estimate. It is evident from the crossing of the two curves that for time constants longer than about one second it may be possible to construct detectors whose sensitivity would be limited by temperature noise, rather than by Johnson noise.

It will be noted that Havens' limit depends on the time constant by the inverse first power of the time constant, rather than by the inverse one half-power. This difference arises because in temperature detectors limited by Johnson noise the coupling coefficient between the thermal and electrical effects is all important, and because in the case of the bolometer, for example, this coupling coefficient depends on the time constant; if, in a given bolometer, the cooling constant is increased with a corresponding decrease in the time constant, the voltage which may be placed across the element is increased, so that the coupling coefficient is also increased.

The Photoelectric Cell

Consider secondly a photoelectric cell with an area of one square millimeter and with a quantum efficiency of one percent for all frequencies greater than 3.0×10^{14} c.p.s., which is the frequency corresponding to a wave-length of one micron. The cell is to have zero quantum efficiency for lower frequencies.

Then, for any signal composed entirely of frequencies greater than 3.0×10^{14} c.p.s., the ratio ϵ_u/ϵ_s^2 which enters into (3.8) has the value

$$\frac{\epsilon_u}{\epsilon_s^2}=100\cdot\frac{15}{4\pi^2}\int_{\xi_0}^{\infty}\frac{\xi^4e^{-\xi}}{(1-e^{-\xi})^2}d\xi \qquad (9.3)$$

by use of Eqs. (4.1) and (4.9). The quantity ξ_0 is defined by

$$\xi_0\equiv h\nu_0/kT, \qquad (9.4)$$

where ν_0 is the cut-off frequency of the detector, taken here to be 3.0×10^{14} c.p.s. Thus one has with $T=300°\text{K}$:

$$\xi_0=48. \qquad (9.5)$$

For values of ξ_0 greater than 48 the exponential factor in the numerator integrand in (9.3) varies so much more rapidly than the other factors that the latter may be considered as constant; one thus has

$$\epsilon_u/\epsilon_s^2=3.85\xi_0^4e^{-\xi_0}\leqq3\times10^{-14}. \qquad (9.6)$$

The square root of this factor must now be appended to (9.1), because this factor was assumed to be unity in deriving this relation. One thus finds

$$H_m\leqq\frac{5\times10^{-19}\text{ watt}}{\tau^{\frac{1}{2}}} \qquad (9.7)$$

for the minimum detectable power of the photoelectric cell in question.

For comparison with this relation, the power represented by one photon per second is 2×10^{-19} watt, when the photons all have a frequency of 3.0×10^{14} c.p.s. Thus the photo-cell in question, with a time constant of one second, would just comfortably be able to detect one photon per second.

Since the factor (9.6) depends very strongly on the cut-off frequency ν_0, changing by a factor of 100 when ν_0 changes by ten percent, it follows rather generally that photo-cells should be able to detect single photons whenever their cut-off wave-lengths are appreciably shorter than one micron.

Experimentally, however, commercial photomultiplier tubes with their peak sensitivities in the visible and with very sharp cut-offs toward the red are not able to detect individual photons unless they are cooled down to liquid-air temperatures.[6] Accordingly, it must be concluded that these tubes contain other sources of noise; the chief source of the additional noise is stated by Engstrom to be thermionic emission from the photo-cathode.

ACKNOWLEDGMENT

The writer is heavily indebted to the pioneer paper on this subject by Milatz and van der

[5] R. J. Havens, J. Opt. Soc. Am. **36**, 355A (1946).

[6] Ralph W. Engstrom, J. Opt. Soc. Am. **37**, 420 (1947).

Velden,[7] for the stimulation which led to work reported here. He is further indebted to Professor John Strong, of Johns Hopkins University, and to Dr. Bruce Billings, of Baird Associates, for helpful discussions about the matters discussed above.

APPENDIX

In order to inter-compare a number of bolometers whose properties have been measured in quite different manners, it is necessary to calculate the performance the bolometers would have if they were operated under a reference condition. The situation is complicated by the fact that with all actual bolometers, whose sensitivity is limited by Johnson noise, it is necessary to specify the frequency response of the amplifier, as well as of the bolometer.

The reference condition selected for the inter-comparison used here is as follows: The bolometer is assumed to be irradiated by a sinusoidal signal whose angular frequency is the reciprocal of the time constant τ of the bolometer. Thus the frequency f is related to τ by

$$f=1/2\pi\tau. \tag{9.2}$$

The amplifier is assumed to be flat at all frequencies, except for a simple RC high frequency cut-off whose time constant, defined by RC, is equal to the time constant of the bolometer. The incident power is measured by half of its peak-to-peak value, and the minimum detectable power is defined by the condition that half of the peak-to-peak signal voltage is to be equal to the r.m.s. noise voltage. (As pointed out in Section VI, however, the same numerical result for the minimum detectable power will be obtained if any measure of the incident power and of the signal voltage is used, provided only that the measures be the same.) The result is to be reduced to an area of one square millimeter on the basis that the minimum detectable power should be proportional to the square root of the area. (This dependence on the area is predicted not only by the thermodynamic theory given here, but also by more detailed theories of the operation of bolometers.)

The reference condition is an optimum one in two respects. First, if one wishes to detect a signal of frequency f and is able to vary the effective cooling constant of the bolometer, the greatest signal voltage is obtained if the cooling constant is adjusted so that the time constant is related to f by Eq. (9.2). Once this adjustment has been made, the greatest signal-to-noise ratio is obtained if the time constant of the high frequency cut-off of the amplifier is adjusted equal to that of the bolometer. The noise equivalent band width of such an amplifier characteristic is $\frac{1}{2}\pi f$.

The writer feels that the reference condition specified above has much to recommend it, since it has only one arbitrary numerical specification (area of one square millimeter) which is not uniquely specified by the behavior of the bolometer itself. Perhaps its greatest weakness is its restriction to bolometers which can be characterized by a single time constant. In any event, however, the evaluation of bolometers with multiple time constants will be essentially more complicated. Accordingly, the writer recommends its use in future reports of bolometer sensitivities, since the computed performance under the reference condition permits immediate comparison with the bolometers listed in Fig. 1.

It will be noted that the reference condition involves testing the bolometer with a sinusoidal signal with a frequency specified by Eq. (9.2), whereas the expression (3.8), which is used in plotting the line representing the thermodynamic limit in Fig. 1, refers to testing with an infinitely long pulse. It is evident that the use of a long pulse instead of the sinusoidal signal would reduce the minimum detectable power by the factor one-half. The use of the sinusoidal signal rather than a long pulse was chosen for the reference condition because the sinusoidal signal is much more closely related to the means of testing actually used. In comparing the experimental points in Fig. 1 with the line representing the thermodynamic limit, however, it should be borne in mind that the points should be lowered by a distance corresponding to a factor of two.

Until such time as bolometers are available whose sensitivity is limited by temperature noise rather than Johnson noise, bolometers of different time constants may be inter-compared on the

[7] J. M. W. Milatz and H. A. van der Velden, "Natural limit of measuring radiation with a bolometer," Physica **10**, 369 (1943).

basis that their minimum detectable power under the reference condition should be inversely proportional to their time constants, as suggested by Havens' limit.

The Baird Bolometers

The slower of the two Baird bolometers is reported† as follows:

$$A = 3.6 \text{ mm}^2, \quad \tau = 50 \text{ milliseconds.}$$

With a square-wave heat signal with a frequency of 10 c.p.s., the electrical response with a wide band (2–5000 c.p.s.) amplifier is 3.0 volts, read on a meter which measures the average absolute value of the voltage and which is calibrated to read the r.m.s. value of a sine wave, per peak-to-peak incident watt. The noise was measured separately in a band 8 c.p.s. wide (noise-equivalent band width) centered at 30 c.p.s.; the meter just described read 3.5×10^{-9} volt.

The observed response is equivalent†† to 2.7 average volts per peak-to-peak watt. If one ignores the small phase shift introduced by the low frequency amplifier cut-off of 2 c.p.s., it is easily computed that the 0.05-second time constant of the bolometer causes the average voltage to be only 0.164 of the peak-to-peak voltage obtained from a very low frequency square-wave signal. Thus at very low frequencies the response of the bolometer is 16.5 volts per watt, with any consistent measure of the power and of the voltage. In the reference condition, the effect of each of the two time constants is to reduce the response by the square root on one-half, so that the response in the reference condition is 8.2 volts per watt.

The measured noise voltage is equivalent††† to 3.96×10^{-9} r.m.s. volt in an 8-c.p.s. band, or to 3.13×10^{-9} r.m.s. volt in a 5-c.p.s. band, the band width corresponding to the reference condition.

It follows directly from the results of the last two paragraphs that the minimum detectable power of the bolometer in the reference condition would be 3.82×10^{-10} watt. Upon division by the square root of 3.6 in order to reduce the result to that for an area of one square millimeter, one obtains the result given in Fig. 1 of 2.01×10^{-10} watt. With a very long pulse instead of the sinusoidal signal, the minimum detectable power would be 1.00×10^{-10} watt.

The faster of the Baird bolometers is reported as follows:

$$A = 0.2 \text{ mm}^2, \quad \tau = 4.1 \text{ milliseconds.}$$

With a square-wave heat signal with a frequency of 10 c.p.s. the electrical response with a wide-band (2–5000 c.p.s.) amplifier is 2.75 volts, read on a meter which measures the average absolute value of the voltage and which is calibrated to read the r.m.s. value of a sine wave, per peak-to-peak incident watt. The noise was measured separately in a band 8 c.p.s. wide centered at 30 c.p.s.; the meter just described read 3.5×10^{-9} volt.

The observed response is equivalent to 2.47 average volts per peak-to-peak watt. Since the ratio of the bolometer time constant to the period of the square wave is the same as for the slower bolometer, the same factor of 0.164 should be used. One finds that at very low frequencies the response of the bolometer is 15.1 volts per watt, with any consistent measure of the power and of the voltage. In the reference condition the response is reduced to 7.6 volts per watt.

The measured noise voltage is equivalent to 3.96×10^{-9} r.m.s. volt in an 8-c.p.s. band, or to 1.09×10^{-8} r.m.s. volt in a 60-c.p.s. band, the band width corresponding to the reference condition.

It follows directly from the results of the last two paragraphs that the minimum detectable power of the bolometer in the reference condition would be 1.43×10^{-9} watt. Upon multiplying by the square root of 5 in order to reduce the result to that for an area of one square millimeter, one obtains the result given in Fig. 1 of 3.20×10^{-9} watt. With a very long pulse instead of the sinusoidal signal, the minimum detectable power would be 1.60×10^{-9} watt.

† Private communication from W. G. Langton, dated April 15, 1947.

†† A meter which measures the average value of a signal, and which is calibrated to read the r.m.s. value of a sine-wave signal, should have its indication multiplied by $2^{\frac{1}{2}}/\pi = 0.9003$ in order to obtain the average voltage.

††† A meter which measures the average value of a signal, and which is calibrated to read the r.m.s. value on a sine-wave signal, should have its indication multiplied by $2/\pi^{\frac{1}{2}} = 1.1284$ in order to obtain the r.m.s. voltage of a noise signal, or of any other voltage with a Gaussian distribution of amplitudes.

The Bell Telephone Laboratories Thermistor Bolometer

The thermistor bolometer is reported‡ as follows:

$$A = 0.6 \text{ mm}^2, \quad \tau = 3 \text{ milliseconds.}$$

A typical quartz-backed thermistor bolometer gives a signal which is equal to the noise when 2×10^{-8} watt of radiation falls on the bolometer for 0.003 second for a 30-c.p.s. pass band.

If one assumes that the "signal equal to the noise" means that the peak signal voltage is equal to the r.m.s. noise voltage, it is still not possible to make an exact reduction to the reference condition without making a detailed calculation of the transient wave form, and this cannot be done without more detailed information about the phase and gain characteristics of the amplifier. Accordingly, only an approximation reduction will be made.

It is estimated that if the pulse of radiation were infinitely prolonged, and if the amplifier passed direct current, the signal voltage would be roughly twice that actually measured. In order to obtain the minimum detectable power under the reference condition, the change in both the signal voltage and the noise voltage will now be computed for an amplifier which has the same maximum gain as that actually used, and which is flat except for a high frequency cut-off with a time constant of 0.003 second, and therefore at a frequency of 53.1 c.p.s. Under this reference condition the signal voltage will be one-half of its low frequency value, which change just annuls the increase estimated in the first sentence. Under this reference condition, the noise-equivalent band width is 83.3 c.p.s. instead of 30 c.p.s., so that the noise voltage is increased by the factor 1.67. Accordingly, the signal-to-noise ratio in the reference condition is reduced by the reciprocal of 1.67, so that the minimum detectable power under the reference condition is computed to be 3.34×10^{-8} watt. Upon dividing by the square root of 0.6 in order to reduce the result to that for an area of one square millimeter, one obtains the result given in Fig. 1 of 4.31×10^{-8} watt. With a very long pulse instead of the sinusoidal signal, the minimum detectable power would be 2.15×10^{-8} watt.

‡ W. H. Brattain and J. A. Becker, J. Opt. Soc. Am. **36**, 354A (1946).

The Polaroid Bolometer

The Polaroid metal bolometer is reported‡‡ as follows:

$$A = 6 \text{ mm}^2, \quad \tau = 4 \text{ milliseconds.}$$

With an amplifier with maximum gain at 30 c.p.s. and with a noise-equivalent band width of 100 c.p.s., the r.m.s. signal voltage is equal to the r.m.s. noise voltage when the r.m.s. signal power incident upon the bolometer at 30 c.p.s. is 3.3×10^{-8} watt.

Under the stated conditions, the minimum detectable power is thus 3.3×10^{-8} watt. In order to obtain the minimum detectable power under the reference condition, the change in both the signal voltage and the noise voltage will now be computed for an amplifier which has the same maximum gain as that actually used, and which is flat except for a high frequency cut-off with a time constant of 0.004 second, and therefore at a frequency of 39.8 c.p.s. Under this reference condition the frequency of the signal is 39.8 c.p.s. instead of 30 c.p.s., so that the response of the bolometer itself is reduced by the factor 0.882, and the response of the amplifier is reduced by the square root of one-half, so that the signal voltage is reduced by the factor 0.624. Under the reference condition the noise-equivalent band width is 62.5 c.p.s. instead of 100 c.p.s., so that the noise voltage is reduced by the factor 0.790. Accordingly, the signal-to-noise ratio in the reference condition is reduced by the factor 0.79, so that the minimum detectable power is raised to 4.18×10^{-8} watt. Upon dividing by the square root of 6 in order to reduce the result to that for an area of one square millimeter, one obtains the result given in Fig. 1 of 1.70×10^{-8} watt. With a very long pulse instead of the sinusoidal signal, the minimum detectable power would be 8.5×10^{-9} watt.

‡‡ B. H. Billings, W. L. Hyde, and E. E. Barr, J. Opt. Soc. Am. **37**, 123 (1947).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 11 NOVEMBER, 1947

Red and Infra-Red Filters for the Study of the Herschel Effect

HELMER BÄCKSTRÖM AND JOHAN BASBERG
Photographic Institute of the Royal Institute of Technology, Stockholm, Sweden
(Received July 7, 1947)

The contradictory literature about the Herschel effect can largely be explained by the fact that sufficiently safe filters are not used for the red exposure. The part of the actinic light that may pass the red filter without complications is extremely small. A few red filters and tests made with them are described, and further conditions for good results with Herschel filters are mentioned.

1. INTRODUCTION

IN the well-known work of W. Clark,[1] the author gives a brief summary of the literature concerning the Herschel effect. After referring to the observations of Draper, Herschel, Lerebour, and Claudet, concerning the Herschel effect on daguerreotype plates, and the analogous observations of Waterhouse and Abney on collodion plates, he states that the reports on gelatine plates are most contradictory. Concerning the above mentioned Clark states: "It has often been stated, even by the highest authorities, that the effect is not observed in the case of ordinary plates unless they are bathed in dye solutions before exposure. Some investigators have found it impossible to repeat the observations of Waterhouse and Abney, and some considered that the effect was simply one of solarization. Others have claimed that the effect can be obtained with ordinary non-color-sensitive gelatin silver bromide plates. Lüppo-Cramer has even found that some papers show it markedly, particularly if copper salts are present. There have been so many arguments and so many conflicting experimental results, that it would appear that the effect is not general, but is dependent on the quality of the photographic material used and on the conditions under which it is employed. It is certain, however, that if the right material and conditions are employed, very definite reversal of the effects of exposure to short-wave-length light can be produced by exposure to longer wave-lengths."

The variety of conflicting observations, on which the literature[2] concerning the Herschel effect seems to be so rich, might in our opinion depend on the fact that the question of filter has not been sufficiently observed. Naturally, several scientists in this field—for instance Leszynski[3]—have been aware of the risk of unreliable red filters, but this knowledge has certainly not been general.

The reasons for this might be twofold: In the first place the time usually needed by red exposure is a very lengthy one. Hence it follows that the investigators concerned do not want to increase the exposure more through the use of a very dense red filter, and so instead use one as bright as possible. On the other hand, however, a closer consideration shows that by using red filter for the Herschel effect we need a considerably denser filter than in more general photographic work.

According to the modern view[4] the primary moment of the latent image's production is performed through the absorption of active quanta of light by ions of halogen in the ion lattice of the silver bromide crystals. By the Herschel effect, on the other hand, the red or infra-red quanta of light should be absorbed by the silver germs which then liberate electrons. The probability that a red quantum of light should be absorbed by a germ, compared with the probability that for instance a blue quantum should be absorbed by an ion of halogen is about as inconsiderable as is the section area of the

[1] W. Clark, *Photography by Infrared* (John Wiley and Sons, Inc., New York, 1946), Second Edition; the Herschel effect is treated on pages 105–112.

[2] A full survey of the literature concerning the Herschel effect is to be found in Kennetth Mees' book *The Theory of the Photographic Process* (The Macmillan Company, New York, 1942), 1945 Edition, pp. 277–289.

[3] W. Leszynski, Zeits. f. wiss. Phot. **24**, 275 (1926); referred to in reference 2.

[4] R. W. Gurney and N. F. Mott, Proc. Roy. Soc. **A 164**, 151 (1938).

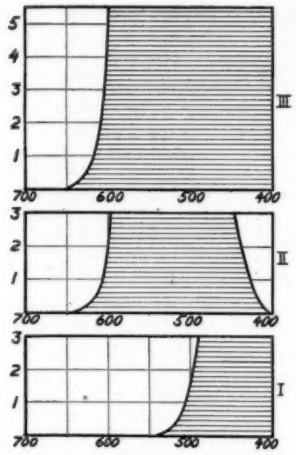

FIG. 1. Density curves for a yellow-filter folie (I), a red-filter-folie (II), and the combination (III) of two red filters (type II) and a yellow filter (type I). The absorption limit is situated at approximately 6000A.

germ compared with the section area of the halogen silver grain in its entirety. In order that a Herschel exposure should cause as great a destructive effect on the latent image as a certain short-wave quantity of light causes production of it, the quantity of light employed in a Herschel exposure must clearly be very much greater. Tollert[5] stated that the requisite quanta of red light was $10^{9\pm1}$ times greater than that required for a similarly great effect caused by short-wave quanta. Even so minute a residual of active light, representing less than 10^{-8} percent of the Herschel light used, must in many cases be considered as a serious source of error in Herschel investigations. Such quantities of residual light generally are so unimportant that they are not taken account of in other photographic work. Even a million times greater quantity of residual light often need not be taken into consideration. We ought perhaps to emphasize, that in this report we particularly refer to the conditions under the "real" Herschel effect, *viz.*, when the photographic layer is not impregnated with any desensitizing dye. But in cases of impregnation more specialized conditions arise, and experts differ as to whether any entirely new moment arises in the mechanism of the reversal action; compare for example the survey of Mees.[2]

[5] H. Tollert, Zeits. f. physik. Chemie **140**, 355 (1929).

2. EXPERIMENTS WITH A RED FILTER (NO. 1); ABSORPTION LIMIT APPROXIMATELY 6000A

To start with, an experiment was made to achieve a red filter, transparent above approximately 6000A, but absorbing shorter wave-lengths by means of a combination of red- and yellow-filter folies. The filter folies had been prepared by bathing fixed-out sheets of Agfa photo-technical film in respective dye solutions. The finished filter consisted of a yellow and two red folies, cemented between glass. The upper absorption limit of the yellow filter (Fig. 1, I) was approximately 5000A and also absorbed the ultraviolet range. The absorption range of the red filters (Fig. 1, II), on the other hand, was between 6000A and 4400A approximately but they were transparent beyond these limits. The density curves were measured by means of a Goldberg's spectrodensograph, which enables measurements up to, and somewhat above density 3. These curves, for each folie taken separately, are shown in diagram 1, I and 1, II. Thus the density of the combination filter can be indicated up to densities about 6 (Fig. 1, III). This combination filter can in general be described as a filter with an absorption limit of approximately 6000A, transparent for the red and infra-red wave-lengths lying immediately above, but opaque for wave-lengths below 6000A. We can conclude from the steep curves that the density within the visible spectrum <6000A is >6 and most likely several units above this figure. The residual light which passes the filter is thus certainly less than 10^{-6} and probably less than 10^{-8} to 10^{-10} for visible light <6000A. It is probably even very small for ultraviolet rays, but in this case there is a degree of uncertainty as we had not access to sufficiently accurate measuring instruments for this range of wave-lengths.

In our investigations of the Herschel effect, which have not yet been published, and for which we tried to find out a suitable filter, we have mostly used a very high intensity of red light. The latter was produced by means of a Bell and Howell 16-mm projector. The metal net protector for stationary projection was removed to secure highest intensity of light. The effect of the projection lamp was 500 watt, and the light was concentrated on a film surface of

only 15×20 mm². Even with the denser filters (Nos. 2 and 3), which will be described below, the Herschel effect of this source of light was considerable. A film pre-exposed to blue light up to a density after development of, for example, $D = 2$, lost nearly the whole of this density after only a few seconds of Herschel exposure. The development took place for 5 minutes in the ordinary DIN-developer at 18°C, brush development, fixation in two baths, 2½ minutes in each.

Let us now return to the lighter filter No. 1. With this we obtained a density of 1.87 over the fog, which was 0.11, with the relatively long red exposure of 30″ on a Gevaert Dipos film which had not been exposed to blue light. It is thus evident that an excessive dose of actinic light passes through this filter.

The nearest explanation for this would be that a small quantity of ultraviolet light passes through the filter and on account of its high actinity overcomes the reversal action of the red light and creates a latent image. To test this, filter No. 1 was combined with a very dense esculin filter, not less than 20 g per m². It is presumed that such an esculin filter must be extremely absorbent of ultraviolet rays; the normal density of esculin is 2 g per m² or less.[6,7,8] When the above experiment was repeated with this combination, we again obtained the considerable density of 1.88 over the fog, i.e., identically the same value as without the esculin filter within the given limit of measurements. Thus it is evident that the fogging effect of the red filter was not caused by short-wave radiation, but also that light waves above 6000A must have a noticeable actinic effect. This fact is also established by other experiments (see next paragraph).

3. THE ABSORPTION LIMIT OF THE RED FILTER

It is interesting to discuss the experimental results mentioned in the preceding paragraph in connection with Hilsch and Pohl's[9] well-known diagram of absorption bands in single large silver bromide crystals; the diagram is given below (Fig. 2). The curve a' in the diagram represents the original absorption range of the silver bromide; the light energy herewith absorbed is largely used to produce a latent image. Already in the blue-green band the absorption and, consequently the sensitivity, decreased to such an extent that the part played by these wave-lengths in the production of the latent image becomes negligible compared with the combined effect of the ultraviolet, violet, and blue light.

Curve b represents the Herschel range, whose absorption band first appears after a latent image ("germs") has been obtained in the crystal by means of a pre-exposure in the absorption range a'. In this case maximum absorption is attained with relatively long wave-lengths, as shown here approximately 7000A. Other experiments, for example Bartelt and Klug's[10] show that with increasing Herschel exposure its action shifts towards longer wave-lengths (e.g. 8000A or more).

The curves a' and b are not, however, on the same scale. As already stated in Section 1 the red exposure must be of quite a different dimension, possibly $10^{9\pm1}$ times greater than the blue exposure in order that the effect of the former could be comparable.[5] In Hilsch and Pohl's diagram this has been taken into account, and consequently the curve a, which corresponds with a part of a', has been enlarged to a scale 10^6 times greater. If firstly we presume that the scale in the curves a and b has been well chosen and, further, that the sensitivity values correspond

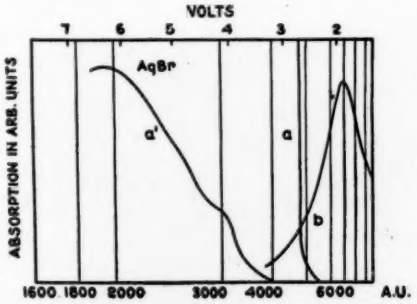

FIG. 2. The absorption bands of silver bromide (Hilsch and Pohl).

[6] Helmer Bäckström and Robert Johansson, Phot. Industrie **34**, 1189 (1936).
[7] Helmer Bäckström and Robert Johansson, Zeits. wiss. Phot. **36**, 194 (1937).
[8] Helmer Bäckström and Robert Johansson, Zeits. wiss. Phot. **38**, 213 (1939).
[9] R. Hilsch and R. W. Pohl, Zeits. f. Physik **64**, 606 (1930).
[10] O. Bartelt and H. Klug, Zeits. f. Physik **89**, 779 (1934); compare also Fig. 93 in reference 2, p. 284.

with the absorption values, then we see that the silver-forming action extends up to about 5500A and the reversal action possibly down to 3000A. For every wave-length in this field there is consequently both a constructive and a reversal action, and this applies even to wave-lengths far beyond the limits given, although each tendency in this case is so insignificant that it does not appear in the diagram. With a wave-length of approximately 4800A the curves should intersect; above this the Herschel effect should be predominant, while below the latent-image formation is the stronger. From what has already been said it is evident that the curve a' should not only have been enlarged to a scale of 10^6 times greater but possibly even 1000 times greater still. In this case the point of intersection of the new a curve and the b curve would be shifted still further in the direction of longer wave-lengths, and the limit of the predominance of the constructive tendency over the reversal effect is now found to be above 6000A, in accordance with the experiment described in the above paragraph.

In this connection we must also point out *that the limit is also shifted in accordance with the intensity of the active light.* The lower this is, the more the silver-forming tendency overcomes the reversal effect. This follows from theoretical reasons: At low intensity of the active light the liberated electrons, and consequently the silver formed, are concentrated possibly only on a single germ in the grain, but at a high intensity the silver formed is spread over many germs. This has been experimentally shown by Berg, Marriage, and Stevens.[11] The consequences of this appear to us to be: firstly, that the effectiveness of *active* light is less (the reciprocity-law failure at high intensity, compare reference 4), secondly that the reversal tendency must be greater, because the red quanta, which could be absorbed, must increase in proportion with the number of the germs (more strictly the total of their sectional area).

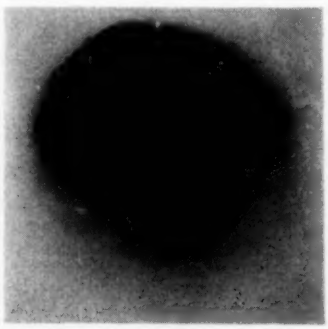

FIG. 3. Spot of light with an area of 15×20 square mm projected on a film by a 16-mm projector.

The two experiments with filter No. 1 (absorption limit 6000A) described in the above paragraph confirm this experimentally. A reproduction of the film in the first of these experiments is shown in Fig. 3. As already stated, in these experiments the very strong light from a 16-mm film projector was concentrated principally on a 15×20 mm² area of the film. It is evident that even outside the limit of this area a certain radiation took place, so that the intensity of light diminished the further it went sidewards. On developing the film it was seen that the density on the central, rectangular area, which had received the strongest light, was 1.88, respectively, 1.87 over the fog. But just around this area the film showed a much higher degree of density which, however, diminished with the distance beyond the area in question. The conditions can be described as follows: The tendency of the faint light furthest away is predominantly constructive, but because of the low degree of intensity the density is moderate (about 1.2). The higher the intensity the greater the density up to a certain limit, when maximum density is obtained. Above this point the density again diminished in spite of the increasing intensity of light. This is due to the fact that the Herschel effect now increases faster than the *constructive* tendencies.

Experiments of a different kind also point in this direction. If a blue pre-exposure took place through a wedge in contact with the film, then naturally, after developing, a certain series of density values on the film was obtained. If, however, before developing the exposed film was subjected to a new exposure, but this time to red (filter No. 1) equally over the whole film area, it was found that the density had increased on the areas with low blue exposure and low density

[11] W. F. Berg, A. Marriage, and G. W. W. Stevens, Phot. J. **81**, 413 (1941); J. Opt. Soc. Am. **31**, 385 (1941).

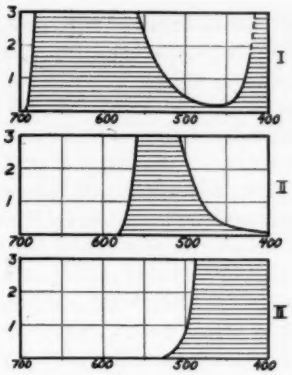

FIG. 4. Density curves for the separate filter elements which combined give the curve in Fig. 5. The (I) refers to Patentblau (3.2 g/sq. m), (II) refers to rose bengale (1.0 g/sq. m), and (III) refers to tartrazin (4.0 g/sq. m) and esculin (10.0 g/sq. m).

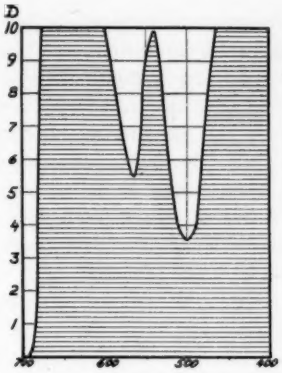

FIG. 5. Density curve for filter 2.

values, but on the contrary it had considerably decreased in the denser areas. The same red exposure can thus, according to the prevailing conditions, give rise to either increased *formation* of the latent image or to its weakening by means of the Herschel effect. These latter experiments agree with the results obtained by Gorokhovskii and Shestakoff's[12] experiments, which were conducted in a different way.

To summarize the results obtained in this section, it can be stated that the critical wavelength limit is relatively high—well over 6000A—and that it is not definitely fixed but varies *inter alia* with the intensity of the radiation.

4. RED FILTER WITH ABSORPTION LIMIT OF 6900A

Later we made some experiments with a filter, prepared already in 1927 by docent J. Hertzberg of this institution, and hereinafter termed No. 2. This contained Patentblau 3.2 g per m², rose bengale 1.0 g per m², tartrazin 4.0 g per m², and esculin 10.0 g per m². The absorption limit of this filter is 6900A. This filter was constructed by Hertzberg for ultrared photography; consequently he did not try to achieve the great density in the absorption area, which was found necessary for Herschel experiments (see Section 1). Since he did not make any absorption measurements for this filter but only tested its appropriate composition with the aid of a small, but very effective, spectroscope, we have consequently examined the absorption characteristics of the filter in a similar way as we did for filter 1 (Section 2).

The curves of density for the single-filter elements are shown in Fig. 4, where I intends 3.2 g Patentblau per m², II rose bengale 1.0 g per m², and III 4.0 g tartrazin per m² plus 10.0 g esculin per m². Density values up to 3.0 have been directly measured by help of a Goldberg's spectrodensograph. In the combined curve (Fig. 5) the density values are supposed to be rather exact up to density 10, as the values have been found through addition of the measured values from diluted solutions.

The curve of density for the combined filter is shown in Fig. 5. Actually the weakest point is to be found at about 5000A where a narrow transparent band is to be found; density here is 3.6 only. With our first more qualitative investigations, using the spectroscope with large aperture, the passing green light could not be observed. By repeated experiment, with the filter as well as the spectroscope immediately in front of a 500-watt clear photo-bulb, one was however able to observe a narrow, esmerald-green band.

As we shall at once show, the transmitted green light generally is of very little importance. Partly the unsensitized film material's sensibility is here rather weak and as the band is narrow, the influence of the latent-image forming will

[12] V. N. Gorokhovskii and S. A. Shestakoff, J. phis. Chemie, U.S.S.R. 11, 356 (1938); referred to in reference 2.

become at once very unimportant in comparison to the Herschel effect from the red light, which is practically passing without absorption with a broad band above 6900A, and which at the same time includes the maximum of the Herschel sensibility. Therefore, we have carried out a part of our still unpublished Herschel investigations with some filters of this type, and by this continued control it was proved that there are no such complications whatsoever.

As it is comparatively easy to eliminate the green transparent band we have at the same time composed a still more dense filter (No. 3) by way of doubling the amount of rose bengale. In this way the absorption density rises in the filter's weakest green area to more than 6.0 (see Fig. 6). It is necessary to stress that still more increase of the rose bengale amount by no means narrows the transparence range of the filter in red and infra-red and is therefore possible if wanted. We have, however, never found it necessary to do this, because filter No. 2 seems to be already sufficient.

With filter 2 quite a number of different control tests have been done. To begin with, single film sheets have been blue-exposed in order to make their latent image equivalent to a long series of density values, and furthermore the whole was red-exposed by use of filter 2, and developed. No increase of density happened in the least exposed parts, as it has been the case with filter 1, but increased Herschel effect could be found with increasing primary density. Therefore the Herschel effect is nowhere hidden by any opposite effect. With this experimental series the intensity of the red light was of medium strength, and the exposing time was therefore 8–60 min.

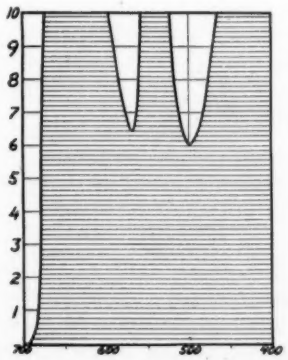

FIG. 6. Density curve for filter 3.

Also experiments with red light of very high intensity have been done (the 16-mm film projector). The time of exposure in this case has commonly been only a few seconds. At a control also 30″ exposure with Gevaert Dipos-film was used. The exposure was done without preexposure with blue light. This truly strong exposure gave, however, only a density of 0.11 over the fog, which in itself was 0.11. Any density maximum at the limits of the projections area—which resulted from experiment with filter No. 1—have not been observed.

The experiment has been repeated with not less than 5 minutes exposure but otherwise under exactly equal conditions. After developing the fully lighted area showed the faint density of 0.13 over the fog, the latter was, as before, 0.11. But at the same time one could observe an extremely fine and somehow darker line of about 0.1-mm breadth around the square.

These experiments show that filter 2 is actually reliable. Experimenting with filter 3 and 30″ exposure the result was a density of 0.14 over the fog, which still was 0.11. Within the exactitude of experiments we can therefore say that filter 3 is equivalent to filter 2.

As filter 3, however, should be expected to have higher security under critical conditions, we want to point out its compound in the same connection, i.e., Patentblau 3.2 g per m^2, rose bengale 2.0 g per m^2, tartrazin 4.0 g per m^2, and esculin 10.0 g per m^2. It is not our intention to state that this filter is better than filters made according to other methods. Naturally one can reach similar results through different combinations of other dyes. It is sufficient for us to stress that filters 2 and 3 have shown themselves trustworthy at Herschel experiments in connection with unsensitized film material.

5. SOME REMARKS ABOUT FLUORESCENCE BY FILTER

Before finishing the discussion of the filter question, it may be important to draw attention to a fact which eventually could result in complications with Herschel experiments. Helmer Bäckström and Robert Johansson of this insti-

tution have shown earlier[6,7,8] that quite frequently filters produce a strong fluorescence. Especially when photographing in mountains, where the sunlight's ultraviolet amount is high, there is a great chance to get a pronounced fog by using such a filter.[6]

With Herschel experiments light sources of comparatively low amount of ultraviolet radiation are used for red exposure. In consequence of what has been said before about the demands of such filters, we see that ultraviolet radiation must be completely eliminated. Therefore, a filter element of strong ultraviolet absorption has to be part of this filter. Such, ultraviolet filters, however, are often strongly fluorescent; i.e., this is the case with esculin.[6,7,8] If such a filter element is met by ultraviolet radiation, the intensity of it decreases when passing through the filter. In reality the emittance of fluorescent light decreases in the same way. Depending upon the color of the fluorescent light and the absorption qualities of the filter element for such radiation, more or less of it is passing through the filter element. If, for instance, the ultraviolet-absorbing layer is colorless, but fluorescent light blue or green, the latter would pass the filter entirely.

Consequently the following is the result: *By use of the Herschel red filter it is of importance which side of it is turned towards the light source and which side is turned towards the film.*

To make this clear we propose the following case. We suppose that the Herschel red filter is composed of two components. One (I) of these is almost completely transparent within the visible spectrum but is completely absorbing every ultraviolet radiation, whereby a visible fluorescent light (for instance blue) is emitted. The other filter component (II), in contrast, is completely opaque for visible radiation of shorter wave-lengths than the limits of absorption in red, but has at the same time a certain transparence of ultraviolet.

If we are putting the filter component I before component II in the way of light, everything will be all right. The ultraviolet radiation will be absorbed in I and as well the direct visible radiation (with exception of red), as fluorescent light from I will be absorbed in II before the light reaches the film.

But if we turn the combined filter then component II absorbs the visible light with shorter wave-lengths than red, but lets the red as well as the ultraviolet radiation pass through. In filter component I the latter is absorbed, but the red radiation is permitted to pass. In addition, there will be the fluorescent light which appears in filter I. The film is hit in this way not only by the expected Herschel light but also by the short-waved fluorescent light.

This example is certainly of difficult construction and in many cases the ultraviolet-absorbing component hinders the transmission of much or perhaps all of the fluorescent light because of dense filter elements. It has been pointed out in Section 1 that extremely small amounts of actinic radiation are already able to exert a detrimental influence; therefore we have good reasons to be interested in the question of fluorescence. Practical examples on the importance of the filter direction have been mentioned in earlier articles by this institution.[7,8]

6. SUMMARY

1. The literature concerning the Herschel effect is filled with obvious contradictions. In the opinion of the authors this depends largely on the fact that the question of filter by red exposure has not been sufficiently observed.

2. A closer investigation shows that even extremely small quantities of actinic light (possibly $<10^{-8}$ percent) could completely distort the result of the Herschel exposure (Section 1).

3. The radiation of the wave-length 6000A is far too actinic to be permitted to exist in noteworthy amounts in this radiation which is passing the red filter. With such radiation we get a strong fog on an unsensitized layer not pre-exposed with another radiation. The absorption limits of the red filter therefore have to be put above 6000A (Section 2).

4. This experimental result can be foreseen from Hilsch and Pohl's classic diagram concerning the absorption bands in exposed silver bromide crystals (Section 3), of which the consequences are discussed in regard to a different question.

5. Radiation of a certain wave-length causes silver-forming as well as reversal tendencies on

the latent image. At longer wave-lengths the latter tendency increases, at shorter, the former increases. At a certain wave-length these tendencies might equalize each other.

6. This utmost wave-length, above which the Herschel effect is dominating over the silver-forming tendency, is not quite fixed. It depends on the intensity of radiation (Section 3).

7. High intensity is favorable to the Herschel effect, low intensity favors the strengthening of the latent image. The higher the intensity, the more is the utmost wave-length shifted downwards (Section 3).

8. The latent image, which is the result of low blue intensity, can be strengthened by help of red radiation which, on the other hand, strongly weakens the latent image produced by higher intensity with blue exposure (Section 3).

9. Direct red exposure of a unsensitized film (without pre-exposure of blue) gives maximal density for a certain intensity. Weaker as well as stronger intensities give less densities (Section 3). The effect shows very clearly, for instance, with a filter of which the absorption limit is lying near 6000A.

10. Red filters with absorption limits near 6900A appear in practical experiments suitable for Herschel exposure. The special effects mentioned above do not appear (Section 4).

11. Advice for such a filter is given (Section 4).

12. If the ultraviolet-absorbing element in the red filter shows fluorescence, then a source of error is able to appear with Herschel experiments (Section 5).

13. This tendency differs in degree, depending on which side of the filter is turned towards the source of light. It will be least disturbing if the ultraviolet-absorbing part is turned from the film towards the source of light (Section 5).

Finally we wish to acknowledge gratefully that the material here published is part of an investigation, concerning the Herschel effect, which is carried out partly by means of public grants to scientific research at the Institute of Technology.

Physics Abstracts

For many years the dues of the American Physical Society have included a subscription to Physics Abstracts, published by the Institute of Electrical Engineers of London. There has been, however, no convenient method by which persons not members of the A.P.S. could receive this publication. The Optical Society of America has now instituted an arrangement under which any O.S.A. member can receive Physics Abstracts at an annual subscription price of $4.50. All correspondence in this connection should be addressed to the American Institute of Physics, 57 East 55th Street, New York 22, New York.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 11 NOVEMBER, 1947

Stray Light Caused by the Presence of a Luminous Body within the Field of an Optical System*

HOWARD S. COLEMAN
The Optical Inspection Laboratory, The Pennsylvania State College, State College, Pennsylvania

(Received June 12, 1947)

The stray light caused by a luminous body within the field of an optical system has been measured for a variety of conditions. Three pairs of telescopic systems and two pairs of photographic systems were studied. Each pair consisted of one coated and one non-coated specimen. The stray-light measurements were made for various angular distances between a black target in a uniformly illuminated surround and a luminous body which could be varied in brightness with respect to the background. These measurements were made for conditions in which the brightness of the luminous body ranged from ten to ten thousand times the brightness of the target. The optical systems were studied under two conditions of alignment with respect to the target. The first was such that no strong ghost image of the luminous body was likely to fall upon the image formed of the target. The second case was such that the strongest ghost image of the luminous body fell upon the image of the target when the angular separation between the target and the luminous body was 2°. The contrast-rendition values were found to range from approximately 90 to 1 percent for the various cases studied. In general there was somewhat less stray light found in the specimens coated with reflection-reducing films than in the non-coated systems. The general trend of the data, however, was found to be the same for all the specimens studied in that contrast rendition started to drop rapidly when angles between the target and the artificial sun became less than a few degrees.

INTRODUCTION

In a recent paper,[1] a general method was described for measuring the stray light in optical systems. This paper deals with the special case of stray light produced by a luminous body either within or near the edge of the field of an optical system under consideration. There are many examples which might be cited in which observations or photographs are taken in which the stray light caused by high lights produce a considerable loss in detail. This loss in detail has been described in terms of the loss in brightness contrast of the image caused by the stray light. In describing this loss, it has been found convenient to introduce a property of an optical system referred to as the "Contrast Rendition." The contrast rendition is a percentage comparison of the contrast of the image formed of an object with the contrast of the object as measured at the entrance window of the optical system under consideration. The contrast rendition, C. R., may be expressed in terms of the image contrast, C_i, and the apparent contrast of the object, C_a, as shown in Eq. (1).

$$\text{C.R.} = (C_i/C_a) \times 100. \quad (1)$$

The contrast rendition of an optical system has been shown to be a function of brightness characteristics of the region surrounding the object under consideration. The angular subtense of the surround is often limited by the subtense of the object under examination. For the purposes of this paper the subtense of the object is selected such that it may be considered an extended object from the point of view of diffraction theory, but small ($\frac{1}{4}$°) with respect to the surround (180°). It is believed that objects of this subtense are of common interest in visual and photographic detection problems.

Because of the numerous ramifications into even the obvious applications of information concerning the magnitude of the stray light produced by luminous bodies with the field of an optical system, the scope of this paper has been limited as much as possible to the methods used and the numerical results obtained. Special consideration is given to the influence of reflection-reducing films on the contrast rendition of common optical systems.

APPARATUS AND PROCEDURE

The apparatus and procedure for making stray-light measurements is essentially the same as reported previously. The only difference in the apparatus is the addition of a luminous body, referred to as "the artificial sun," which can be

* The work presented in this paper was performed, in part, as Task C of Navy Contract NOrd 7958 between the Bureau of Ordnance of the Navy Department and The Pennsylvania State College.

[1] H. S. Coleman, "Stray light in optical systems," J. Opt. Soc. Am. 37, 434 (1947).

moved about in the field with respect to the target under observation. This is shown schematically in Fig. 1. The artificial sun is constructed such that its presence or absence does not appreciably change the brightness of the background against which the target is viewed. The subtenses of the artificial sun and the target are $\frac{1}{4}°$, as measured from the entrance pupil of the optical system under test.

The procedure consisted of making contrast-rendition measurements of the target (a black hole in a 180° uniformly illuminated surround) for various angular separations between the artificial sun and the target. To cover most practical cases, four different brightness ratios of the artificial sun with respect to the background were used. These were such that the brightness of the artificial sun was 10, 100, 1000, and 10,000 times the brightness of the background. Two extreme conditions of alignment of the target and optical system under test were used. The first was such that a ghost image of the artificial sun was likely to fall upon the image of the target formed by the optical system under test for some angular position with respect to the target as the artificial sun was moved across the field. For purposes of comparison, this condition was selected such that the strongest visible ghost image would fall on the image of the target when the artificial sun was separated from the target by 2°. This required the optical systems to be rotated 1° off axis. The second condition of alignment was such that, in general, no strong ghost image of the artificial sun was likely to fall on the image of the target formed by the optical system under test. In this latter case, the optical system under test was adjusted such that the image of the target would fall on the optical axis.

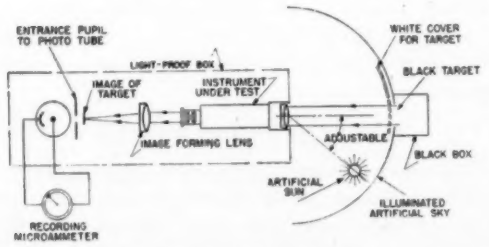

FIG. 1. Schematic diagram of contrast-rendition apparatus and artificial sun.

The optical systems discussed in this paper consisted of three pairs of telescopic systems and two pairs of photographic objectives. Each pair consisted of one specimen having its optical surfaces coated with reflection-reducing films and the other specimen, of the same design but with non-coated surfaces. The telescopic systems consisted of an 8×40×5.2° periscope, a 7×50×7.1° prism erecting binocular, and a 5×35×12.5° lens erecting gunsight. The photographic objectives were of the Altimar and Ic Tessar designs.

THE DATA

The data are presented graphically in Figs. 2 to 21 inclusive.

Figures 2 and 3 show the data on the pair of 8×40×5.2° periscopes with the head prisms used on axis. In this case no strong ghost image of the artificial sun fell upon the image of the target as the artificial sun is moved across the field.

Figures 4 and 5 show the data obtained for the pair of 8×40×5.2° periscopes with the head prisms set at approximately 2.0° depression. At this depression, the strongest ghost visible fell upon the image of the target when the artificial sun was separated from the target by 2°.

Figures 6 and 7 show the data obtained for the right barrels of the pair of 7×50×7.1° binoculars when the image of the target is on the axis of the optical systems. In this particular case, strong ghost image of the artificial sun fell upon the image of the target.

Figures 8 and 9 show the data obtained for the right barrels of the pair of 7×50×7.1° binoculars rotated 1° off axis to the left. For this rotation, a strong ghost image of the artificial sun fell upon the image of the target when the artificial sun was separated from the target by 2°.

Figures 10 and 11 show the data obtained for the pair of 5×35×12.5° gunsights aligned such that the image of the target fell on the optical system's axis. In this case no particularly strong ghost image of the artificial sun fell upon the image of the target as the artificial sun was moved across the field.

Figures 12 and 13 show the data obtained for the pair of 5×35×12.5° gunsights rotated at 1° off axis to the left. For this rotation, a strong ghost image of the artificial sun fell upon the

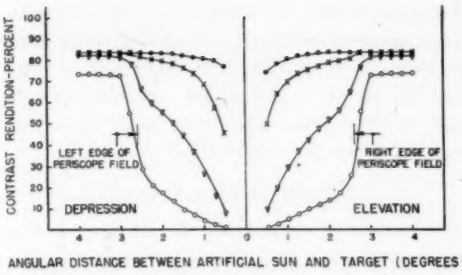

FIG. 2. Axial contrast rendition for various angular distances between the artificial sun and the target for a noncoated 8×40×5.2° periscopic system.

●—Ratio brightness of sun to surround = 10
×—Ratio brightness of sun to surround = 10^2
▽—Ratio brightness of sun to surround = 10^3
○—Ratio brightness of sun to surround = 10^4

image of the target when the artificial sun was 2° from the target.

Figures 14 and 15 show the data obtained for the pair of Altimar objectives aligned with the image of the target falling upon the optical axis. In this case no particularly strong ghost image of the artificial sun fell upon the image of the target.

Figures 16 and 17 show the data obtained for the pair of Altimar objectives rotated such that the target was off axis by 1°. For this rotation, a strong ghost image of the artificial sun fell upon the image of the target when the artificial sun was separated from the target by 2°.

Figures 18 and 19 show the data obtained for the pair of Tessar objectives aligned with the image of the target falling on the optical axis. In this case no particularly strong ghost image of the artificial sun fell on the image of the target.

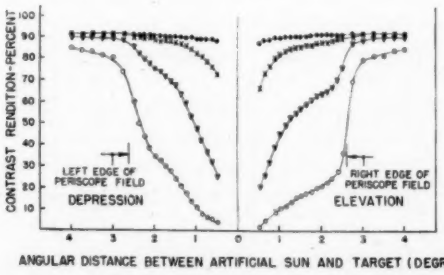

FIG. 3. Axial contrast rendition for various angular distances between the artificial sun and the target for a coated 8×50×5.2° periscopic system.

●—Ratio brightness of sun to surround = 10
×—Ratio brightness of sun to surround = 10^2
▽—Ratio brightness of sun to surround = 10^3
○—Ratio brightness of sun to surround = 10^4

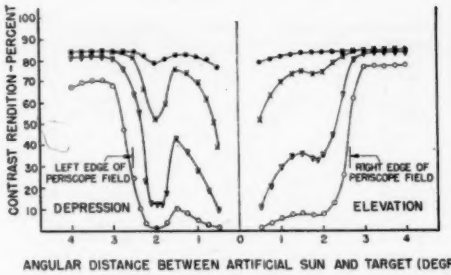

FIG. 4. Axial contrast rendition for various angular distances between the artificial sun and the target for a noncoated 8×40×5.2° periscopic system.

●—Ratio brightness of sun to surround = 10
×—Ratio brightness of sun to surround = 10^2
▽—Ratio brightness of sun to surround = 10^3
○—Ratio brightness of sun to surround = 10^4

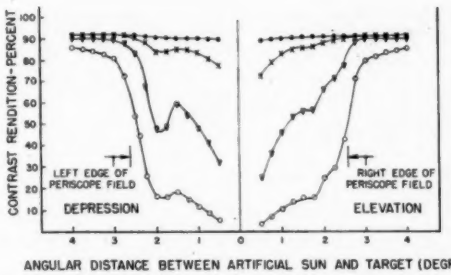

FIG. 5. Axial contrast rendition for various angular distances between the artificial sun and the target for a coated 8×40×5.2° periscopic system.

●—Ratio brightness of sun to surround = 10
×—Ratio brightness of sun to surround = 10^2
▽—Ratio brightness of sun to surround = 10^3
○—Ratio brightness of sun to surround = 10^4

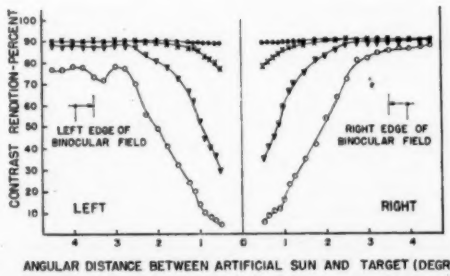

FIG. 6. Axial contrast rendition for various angular distances between the artificial sun and the target for a noncoated 7×50×7.1° prism erecting telescopic system.

●—Ratio brightness of sun to surround = 10
×—Ratio brightness of sun to surround = 10^2
▽—Ratio brightness of sun to surround = 10^3
○—Ratio brightness of sun to surround = 10^4

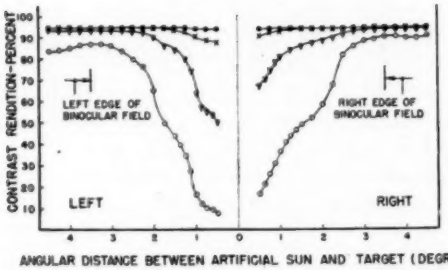

FIG. 7. Axial contrast rendition for various angular distances between the artificial sun and the target for a coated 7×50×7.1° prism erecting telescopic system.

●—Ratio brightness of sun to surround = 10^2
×—Ratio brightness of sun to surround = 10^2
▽—Ratio brightness of sun to surround = 10^3
○—Ratio brightness of sun to surround = 10^4

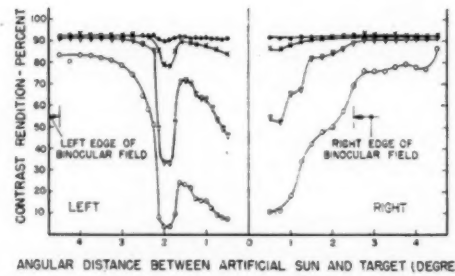
CONTRAST RENDITION - PERCENT
LEFT EDGE OF BINOCULAR FIELD
RIGHT EDGE OF BINOCULAR FIELD
LEFT
RIGHT
ANGULAR DISTANCE BETWEEN ARTIFICIAL SUN AND TARGET (DEGREES)

FIG. 8. Contrast rendition for various angular distances between the artificial sun and the target for a non-coated 7×50×7.1° prism erecting telescopic system rotated 1° off axis to the left.

●—Ratio brightness of sun to surround = 10^2
×—Ratio brightness of sun to surround = 10^2
▽—Ratio brightness of sun to surround = 10^3
○—Ratio brightness of sun to surround = 10^4

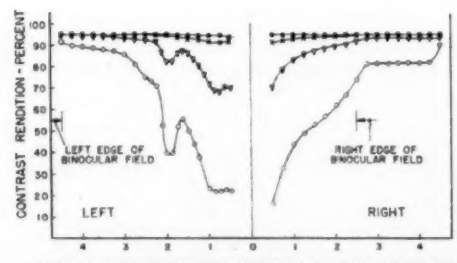
CONTRAST RENDITION - PERCENT
LEFT EDGE OF BINOCULAR FIELD
RIGHT EDGE OF BINOCULAR FIELD
LEFT
RIGHT
ANGULAR DISTANCE BETWEEN ARTIFICIAL SUN AND TARGET (DEGREES)

FIG. 9. Contrast rendition for various angular distances between the artificial sun and the target for a coated 7×50×7.1° prism erecting telescopic system rotated 1° off axis to the left.

●—Ratio brightness of sun to surround = 10^2
×—Ratio brightness of sun to surround = 10^2
▽—Ratio brightness of sun to surround = 10^3
○—Ratio brightness of sun to surround = 10^4

Figures 20 and 21 show the data for the pair of Tessar objectives rotated such that the target was off axis by 1°. For this rotation, a strong ghost image of the artificial sun fell upon the image of the target when the artificial sun was separated from the target by 2°.

DISCUSSION AND CONCLUSIONS

The preceding data indicate that the reduction in contrast caused by stray light from a luminous body within the field of an optical system is of considerable magnitude. The data themselves, however, are not easily analyzed with respect to the causes of stray light. Such an analysis can be made by a visual examination of the image planes and the internal parts of the optical system under consideration. This examination may be accomplished by means of a loupe in the case of an afocal system and by means of a white screen in the case of a focal system. By means of such examinations, the causes of the irregularities in the preceding graphs can be seen (repeated measurements indicate that the irregularities in the graphs are real and are reproducible). The visual examinations reveal that the causes of the individual irregularities may vary widely. Because of this each optical system and set of conditions under which the optical system is being considered must be treated separately. There are, however, a number of general conclusions that can be drawn from this study. These are presented in the following paragraphs:

1. The method used to measure stray light in optical systems is quite versatile and sufficiently

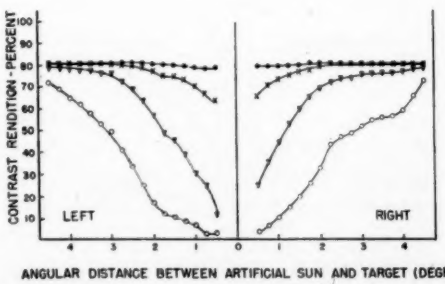

FIG. 10. Axial contrast rendition for various angular distances between the artificial sun and the target for a non-coated 5×35×12.5° lens erecting telescopic system.

●—Ratio brightness of sun to surround = 10^2
×—Ratio brightness of sun to surround = 10^2
▽—Ratio brightness of sun to surround = 10^3
○—Ratio brightness of sun to surround = 10^4

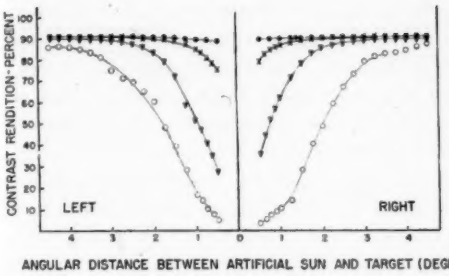

FIG. 11. Axial contrast rendition for various angular distances between the artificial sun and the target for a coated 5×35×12.5° lens erecting telescopic system.

●—Ratio brightness of sun to surround = 10
×—Ratio brightness of sun to surround = 10^2
▽—Ratio brightness of sun to surround = 10^3
○—Ratio brightness of sun to surround = 10^4

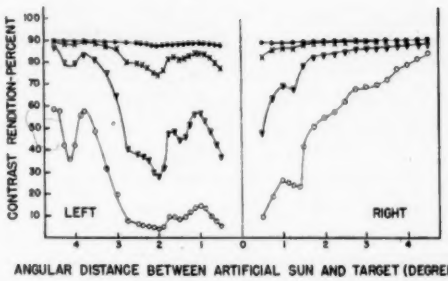

FIG. 12. Contrast rendition for various angular distances between the artificial sun and the target for a non-coated 5×35×12.5° lens erecting telescopic system rotated 1° off axis to the left.

●—Ratio brightness of sun to surround = 10
×—Ratio brightness of sun to surround = 10^2
▽—Ratio brightness of sun to surround = 10^3
○—Ratio brightness of sun to surround = 10^4

precise to study a wide variety of conditions which are of practical interest.

2. The method described to measure the stray light has a certain face validity, in that there is a close correlation between visual or photographic examinations of the causes of stray light and the contrast-rendition data obtained by photoelectric means.

3. It is interesting to note that there is a great deal of similarity in the form of data obtained for the different designs of optical systems considered. It is evident that the true field subtense of an optical system has little relation with respect to the angle between the target and the luminous body at which the contrast-rendition curves start to fall rapidly. This appears to occur from 2° to 5° separation between the target and the luminous body, regardless of the subtense of the field of the optical system under consideration.

4. The use of reflection-reducing films do not eliminate the stray light in the cases reported here. This indicates that a considerable portion of the stray light present is caused by reflection from the mechanical parts. Reflection-reducing films do, however, greatly reduce the stray light in the case of a ghost image falling upon the image of the target.

5. The causes of the irregularities in the graphical data presented here are found to be caused by diffuse and specular reflections from the optical and mechanical parts as well as from a considerable number of ghost images formed when the optical system is used off axis even to a very slight degree. It might be noted that the irregu-

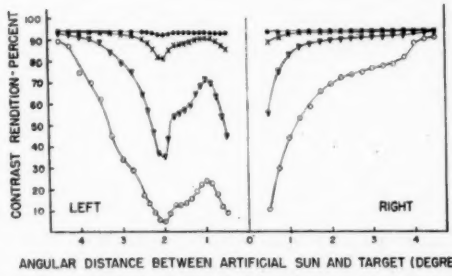

FIG. 13. Contrast rendition for various angular distances between the artificial sun and the target for a coated 5×35×12.5° lens erecting telescopic system rotated 1° off axis to the left.

●—Ratio brightness of sun to surround = 10
×—Ratio brightness of sun to surround = 10^2
▽—Ratio brightness of sun to surround = 10^3
○—Ratio brightness of sun to surround = 10^4

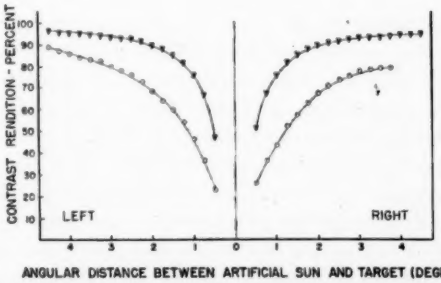

FIG. 14. Axial contrast rendition for various angular distances between the artificial sun and the target for non-coated and coated Altimar photographic objectives at f:5.

Equivalent focal length of objective = 8.25 inches
Ratio (brightness of sun to surround) = 10^4

▽—Coated
○—Non-coated

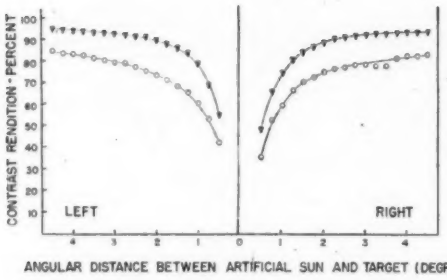

FIG. 15. Axial contrast rendition for various angular distances between the artificial sun and the target for non-coated and coated Altimar photographic objectives at f:11.
Equivalent focal length of objective = 8.25 inches
Ratio (brightness fo sun to surround) = 10^4
▽—Coated
○—Non-coated

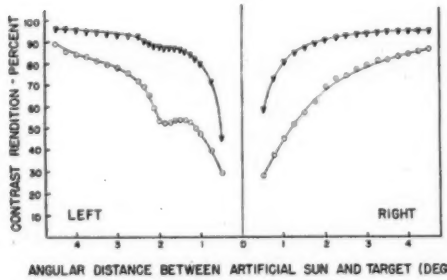

FIG. 16. Contrast rendition for various angular distances between the artificial sun and the target for non-coated and coated Altimar photographic objectives at f:5 rotated 1° off axis to the left.
Equivalent focal length of objective = 8.25 inches
Ratio (brightness of sun to surround) = 10^4
▽—Coated
○—Non-coated

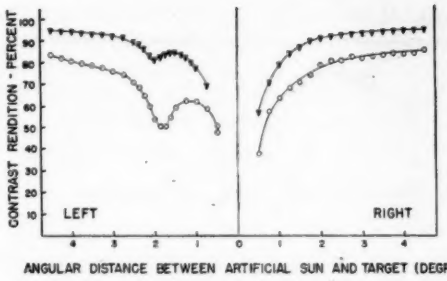

FIG. 17. Contrast rendition for various angular distances between the artificial sun and the target for non-coated and coated Altimar photographic objectives at f:11 rotated 1° off axis to the left.
Equivalent focal length of objective = 8.25 inches
Ratio (brightness of sun to surround) = 10^4
▽—Coated
○—Non-coated

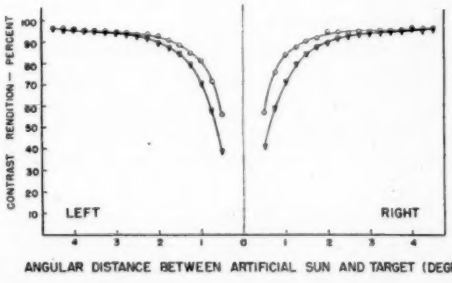

FIG. 18. Axial contrast rendition for various angular distances between the artificial sun and the target for non-coated and coated Tessar photographic objectives at f:4.5.
Equivalent focal length of objective = 8.25 inches
Ratio (brightness of sun to surround) = 10^4
▽—Coated
○—Non-coated

larities caused by these reflections do not always occur symmetrically with respect to either the optical or mechanical axes. This is a result of the non-uniformity in the internal surface treatment of the mechanical parts and of slight differences in the dimensions of the mechanical parts. As an example of the latter, the minimum occurring in the contrast-rendition data at 4° to the left in Fig. 12 is caused by a diffuse reflection from the ground edge of one of the erecting prisms which had not been properly covered by the metal shield provided for that purpose. This same reflection was not found in the case of other non-coated systems of the same design.

6. The axial contrast-rendition values are found to be fairly symmetrical and are generally higher than the off-axis measurements. This indicates that observations or photographs taken of objects near bright luminous regions should be such that the image of the object falls on the optical axis of the system being used.

7. In the case of the Altimar photographic objectives, it is of interest to note that the contrast rendition is higher for the full aperture than at a reduced aperture. This indicates that the image-forming light is reduced more rapidly than the light from the ghost images as the aperture is reduced.

8. The most prevalent causes of ghost images in telescopic systems are the steep curves of the eyepieces and the flat surfaces of prisms, windows, and reticles. The results of experiments not

described here indicate that the use of reflection-reducing films are particularly effective in these cases. This suggests the effective use of reflection-reducing films on haze filters, eye glasses, and windshields (particularly for use at night).

9. The preceding data indicate that the use of ordinary lens shades would not be a practical means of reducing the stray light for natural conditions simulated in this study. The possibility is suggested, however, that an adjustable "image shade" might be developed which would eliminate any desired portion of the field (such as near the image of a bright luminous object).

10. It is seen from the preceding data that there are two separate sources of stray light causing an imperfection in the contrast rendition of optical systems. The first of these is the uniform surround, and the second is the luminous body. Since the stray light is rapidly diminished as the angle between the target and the luminous body is increased beyond a few degrees, contrast rendition remains constant for greater angles, regardless of the presence of the luminous body. This indicates how close observations or photographs can be taken to a bright luminous region with a given reduction in contrast.

11. Figures 18 to 21 inclusive, showing data for the Tessar objective, illustrate a case in which reflection-reducing films do not show a general reduction of stray light. In fact, the data indicate that for some conditions the non-coated specimen had less total stray light than the coated specimen. This can be explained on the basis that the stray light is contributed by scattering from both the optical and mechanical surfaces and that there may be a slight predominance of light scattered from the mechanical parts. It is likely that the reflection-reducing films may transmit more total light to be scattered by the mechanical parts than the non-coated specimen. This does not mean, however, that the amount of image-forming light is greater in the non-coated specimen. In fact, there is little doubt that the image ratio of forming light to the stray light is greater in the coated specimen.

12. The amounts of stray light implied from the contrast-rendition values reported here are conservative for two reasons. The first of these is inherent in the method used, since by definition the image contrast cannot be reduced below zero.

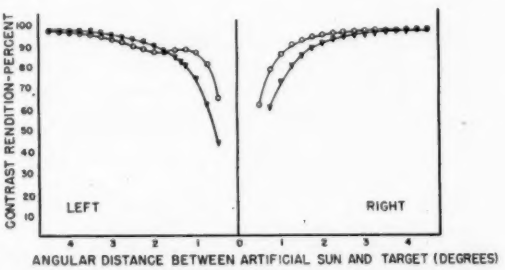

FIG. 20. Contrast rendition for various angular distances between the artificial sun and the target for non-coated and coated Tessar photographic objectives at f:45 rotated 1° off axis to the left.

Equivalent focal length of objective = 8.25 inches
Ratio (brightness of sun to surround) = 10^4

▽—Coated
○—Non-coated

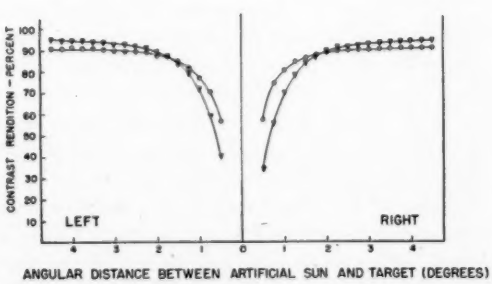

FIG. 19. Axial contrast rendition for various angular distances between the artificial sun and the target for non-coated and coated Tessar photographic objectives at f:11.

Equivalent focal length of objective = 8.25 inches
Ratio (brightness of sun to surround) = 10^4

▽—Coated
○—Non-coated

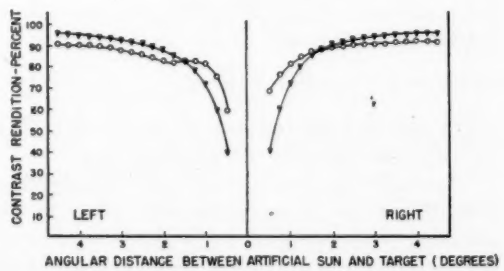

FIG. 21. Contrast rendition for various angular distances between the artificial sun and the target for non-coated and coated Tessar photographic objectives at f:11 rotated 1° off axis to the left.

Equivalent focal length of objective = 8.25 inches
Ratio (brightness of sun to surround) = 10^4

▽—Coated
○—Non-coated

This holds even in the case of a ghost image falling on the image of the target that is many times brighter than the image of the background. The second reason is that there is no account taken of the stray light introduced by the observer's eye or the photographic plate. Such stray light would further reduce the contrast for the conditions discussed in this experiment.

ACKNOWLEDGMENT

The author is indebted to Dr. M. F. Coleman, Mr. S. W. Harding, and Mrs. A. E. Lerew, of the Optical Inspection Laboratory, for invaluable assistance in this study. The photographic objectives were made available through the courtesy of Dr. A. F. Turner of the Bausch and Lomb Optical Company.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 11 NOVEMBER, 1947

Illumination in the Focal Plane of a Camera Lens

EDWARD P. CLANCY
Mount Holyoke College, South Hadley, Massachusetts

(Received January 8, 1947)

Direct photoelectric measurement has been made of illumination in the focal planes of typical photographic objectives. The results in many cases differ seriously from what might be expected simply on the basis of the $\cos^4$ law plus vignetting. The deviation takes the form of an unexpectedly large rate of decrease of illumination with respect to distance from the axis, and may be so great as to have important consequences on work in color photography and photographic photometry. The effect is apparently inherent in the design of almost any photographic objective.

IT is well known that the illumination received by various areas of a photographic film depends considerably on the location of the areas in the focal plane. Consider an ideal thin lens, in a mounting of negligible thickness, placed so that its field of view is of uniform brightness. It can then easily be shown[1] that if E_p is the illumination at a point P in the focal plane, and E_0 is the illumination at the point where the focal plane intersects the lens axis, then

$$E_p = E_0 \cos^4\theta,$$

where θ is the angle between the axis and the line which joins P and the center of the lens.

This effect alone can reduce the illumination near the edges of a photographic plate in an ordinary camera roughly to 60 percent of that at the center, and in the case of wide-angle lenses to 25 percent or less. In an actual lens, however, vignetting—the interception of part of the light by the lens mounting when θ exceeds a certain value—is an added factor in the decrease in illumination. The magnitude of this factor depends greatly on the setting of the diaphragm.

Benford[1] has photographed a camera lens from various angles, observed thereby the effective aperture, and computed the illumination over the focal plane for various diaphragm settings. Washer,[2] in a study of wide-angle airplane-camera lenses, has used a similar technique.

In the course of recent work undertaken by the writer, it has seemed desirable to make a direct photoelectric measurement of illumination in the focal planes of various camera lenses. The results in many cases differ considerably from what might be expected simply on the basis of $\cos^4\theta$ plus vignetting. The deviations are always in the direction of less illumination; i.e., the rate of decrease with respect to θ is unexpectedly large—so large, in fact, as to have serious bearing on work in photographic photometry and color photography.

Typical anastigmatic photographic objectives

[1] Frank Benford, "Note on illumination characteristics of the camera lens," J. Opt. Soc. Am. **27**, 286 (1937).

[2] F. E. Washer, "Characteristics of wide-angle airplane-camera lenses," J. Research Nat. Bur. of Stand. **29**, 233 (1942).

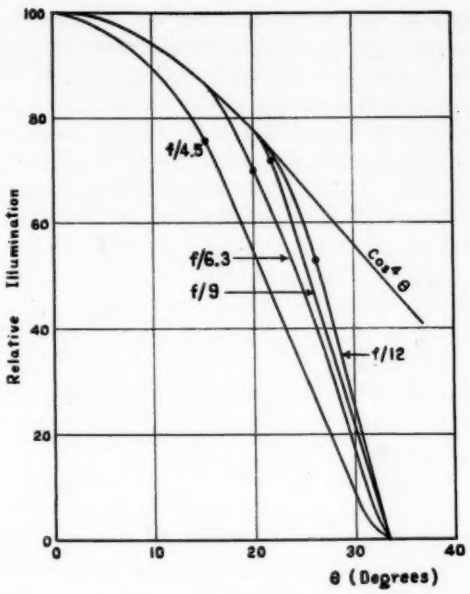

FIG. 1. Illumination in the focal plane of a camera lens. $F=50$ mm. The dots show the values of θ at which vignetting begins with various diaphragm settings.

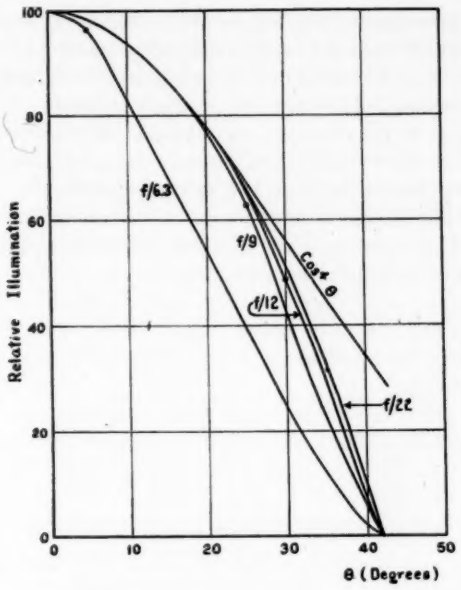

FIG. 2. Illumination in the focal plane of a camera lens. $F=125$ mm.

were used for the measurements. The lens under test was illuminated by collimated light from a tungsten source at 2850°K. Behind the lens was placed a photo-emissive cell, at a distance such that all the light which had passed through the lens fell on the cell's surface. This light was measured as a function of θ by rotating the lens about a vertical axis.

Two correction factors must be added to convert the cell's output into relative illumination: (1) Since the distance from the lens to a point in the focal plane varies as $1/\cos\theta$, and since the illumination, considering the lens aperture as the source of light, varies inversely as the square of this distance, it is necessary to introduce a multiplication factor of $\cos^2\theta$. (2) The light strikes the focal plane at an angle θ; the illumination is therefore reduced by the factor $\cos\theta$.

If the output of the cell be called $f(\theta)$, then the relative illumination E_p/E_0 is proportional to $f(\theta)\cos^3\theta$.

Figures 1, 2, and 3 show the results for lenses of focal length 50 mm, 125 mm, and 230 mm, respectively. In each figure the curve of $\cos^4\theta$ has been drawn in for reference. The large dots on the

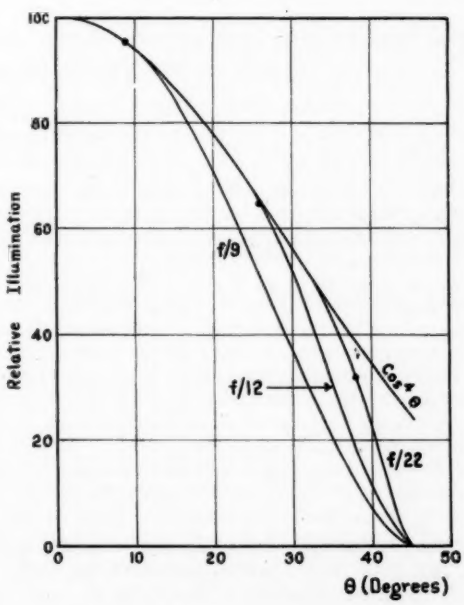

FIG. 3. Illumination in the focal plane of a camera lens. $F=230$ mm. The first dot shows the value of θ at which vignetting begins at f:9.

curves denote the values of θ at which vignetting begins with the various diaphragm settings.

It is obvious at once that in many cases there is unexpected reduction in light intensity.* In Fig. 1, for example, considerable decrease from the $\cos^4\theta$ curve has already occurred, at all diaphragm settings, before vignetting begins. At f:12, the reduction amounts to about 20 percent of the expected value of illumination at the point where vignetting first occurs.

In Fig. 2, the effect is present to a slight degree at f:6.3, and increases steadily with decreasing aperture, becoming large (about 30 percent) at f:22.

The lens of Fig. 3 is least anomalous in its behavior. Here the effect is absent at f:9 and f:12. At f:22, however, the reduction is 18 percent.

Introduction of various filters in the light path showed, within the accuracy of measurement, that the effect is not a function of wave-length in the visible and near infra-red regions. It was supposed that part of the effect, at least, might be due to loss of light in multiple reflections within the lens components. Measurements were therefore taken on two lenses, identical except that one was coated and the other was not. These lenses were obtained through the kindness of the Bausch and Lomb Optical Company. Suitable filters were used to restrict the incident light to the range of wave-lengths in which the coating might be most effective. Tests at various apertures, however, showed no discernible difference in the magnitude of the effect between the coated and uncoated lenses.

It is apparent, therefore, that this unexpectedly large variation in illumination is inherent in the design of almost any photographic objective, although its magnitude may vary considerably from one lens to another.

* F. E. Washer, see reference 2, found a similar phenomenon in two of the airplane-camera lenses which he tested by the indirect method already described.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 11 NOVEMBER, 1947

The Two-Quanta Explanation of the Dependence of the Threshold Values and Visual Acuity on the Visual Angle and the Time of Observation

M. A. Bouman and H. A. van der Velden
Physical Institute of the University of Utrecht, The Netherlands

As shown by van der Velden, for the observation of a short and small light flash it is necessary that at least two quanta of light be effectively absorbed by the visual purple within a time τ (about 0.02 sec.) and within an area corresponding to a visual angle D (about 10′): the two-quanta hypothesis, by which the laws of Ricco, Piper, and Talbot were explained. In the present paper the theoretical foundation is recapitulated and further experiments are described.

The extensive experimental results in the present paper are completely covered by the two-quanta hypothesis as far as regards the flash time $t<3\tau$ or the visual angle of the light spot $d<2D$.

It is shown, that for the simultaneous occurrence of long flashes and large visual angles, considerable deviations from the theoretical curves derived from the two-quanta hypothesis occur.

From these deviations we conclude that T seconds (about 3τ) after the absorption of light quanta the condition of the retina in the neighborhood of these absorptions (within about $3D$) is such as to decrease the chance of observation of a subsequent pair of absorbed quanta.

Measurements of the visual acuity are found to be in agreement with the two-quanta case.

The experimental results of Hecht and Pirenne and the paper by De Vries are discussed, as is also the note of C. Peyrou and H. Piatier about experiments similar to those of van der Velden.

1. THE EXPERIMENTS OF VAN DER VELDEN AND THE THEORETICAL FOUNDATION OF THE TWO-QUANTA EXPLANATION

As a consequence of the quantum character of light, fluctuating results are obtained when determining threshold values.

In 1944 v.d. Velden[1,2] developed three independent approaches to the establishing of the number of quanta effectively absorbed in the

[1] H. A. van der Velden, Physica 11, 179 (1944).
[2] H. A. van der Velden, Ophthalmologica 111, 321 (1946).

visual purple, necessary for the light perception, independent of the loss of light caused by reflection and absorption in the various parts of the eye, so that the absolute values of the energies employed are of no importance. These three methods were based on a determination of the type of relation between the chance of observation of a flash and the average number of quanta of this flash for various visual angles and various times of observation.

The eye of the observer was completely dark-adapted. The spot of the retina, where the flash was received, was about 7° nasal from the fovea. For the experiments in the present paper we used the arrangement already described in the papers mentioned. We shall give in 3 extensive information about the determination of the energies employed.

The conclusion drawn from the experimental results was the two-quanta explanation: a rod reacts on the absorption of one quantum and an impulse will be transmitted to its nerve connection. When two rods give such an impulse, a light impression will result when the two impulses are given by rods lying within an area corresponding to a visual angle D (about 10′) and following each other within a time τ (about 0.02 sec.).

We represent and discuss the three experiments by which the two-quanta explanation was checked under A, B, and C.

A. Determination of the Type of Relation Between the Chance of Observation of a Flash and the Average Number of Quanta in This Flash for Very Small Visual Angles and Short Flash Times

Let k be the number of quanta which must be actively absorbed by the visual purple in order to perceive light, f the fraction effectively absorbed in the visual purple of the total number of quanta, incident on the eye, $\bar{N}$ the average number of quanta per flash, and $W(\bar{N}, k)$ the chance of observation of the flash. Then, according to the Poisson distribution:

$$W(\bar{N}, k) = 1 - \sum_{s=0}^{k-1} [(f\bar{N})^s/s!] \exp(-f\bar{N}). \quad (1)$$

Plotting $W(\bar{N}, k)$ against $\log\bar{N}$, the shape of this curve is independent of the value of f, but different for the various cases $k=1, 2, 3$, etc., so that both f and k can be determined by measuring the chance of observations as a function of $\bar{N}$. These curves are given in the papers of van der Velden[1,2] and of Hecht.[3]

Supposing them to agree with the case $k=2$, f was for three test persons found by van der Velden to be equal to 28 percent, 9 percent and 7 percent. The shape of the curves pointed out that at least two quanta are necessary for the light perception.

In spite of the great differences in the values of f all experimental curves coincided best for the case $k=2$.

B. Determination of the Dependence of the Average Number of Quanta for a Fixed Chance of Observation on the Duration of the Flash for a Small Visual Angle

As the experiments concerning $W(\bar{N}, k)$ could reasonably well be described by the hypothesis that two quanta are necessary for a light perception, it is obvious that one has to expect a dependence of the average number of quanta in the flash $\bar{N}$, necessary for a certain chance of observation, on the duration of the flash t. The retinal condition existing after the absorption of the first quantum will have a finite lifetime τ. If the second quantum is absorbed after this time there will be no light impression at all. For the time of observation $t>\tau$ an extra condition was introduced in formula (1) for $k=2$, expressing the necessity that during the flash at least two quanta follow each other within τ.

Let $W(\bar{N}, t)$ be the chance of observation of the flash of t seconds

$$W(\bar{N}, t) = 1 - \exp(-f\bar{N}) - f\bar{N}\exp(-f\bar{N}) - \sum_{s=2}^{m+1} P_s(\bar{N}, t); \quad m\tau < t \leqq (m+1)\tau$$

$P_s(\bar{N}, t)$ is the chance of no light perception when s quanta are absorbed. $P_s(\bar{N}, t)=0$ when $s>m+1$, as in that case there will of course be at least one couple of quanta absorbed within τ. The chance for absorptions at $t_1, t_2, t_3, \cdots, t_s$ of the s

[3] S. Hecht, J. Opt. Soc. Am. **32**, 42 (1942).

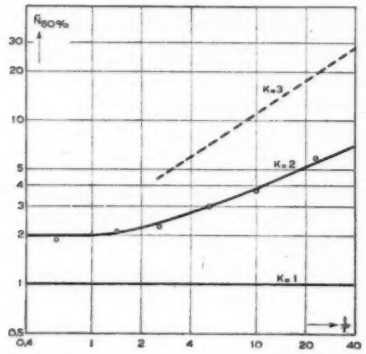

FIG. 1. The theoretical number of quanta, necessary for a chance of observation of 60 percent as a function of t, the duration of the flash, for the 1, 2, and 3 quanta hypothesis. The experimental results for $d \gg D$ from Fig. 7a reduced to $f=1$ are presented.

quanta when the flash starts at $t=0$:

$$\exp[-f\bar{N}/t\cdot t_1]\cdot f\bar{N}/t\cdot dt_1 \times\exp[-f\bar{N}/t\cdot(t_2-t_1)]f\bar{N}/t\cdot dt_2 \times\exp[-f\bar{N}/t\cdot(t_s-t_{s-1})]f\bar{N}/t\cdot dt_s \times\exp[-f\bar{N}/t\cdot(t-t_s)].$$

As for $P_s(\bar{N}, t)$ the interval between two absorptions must always exceed τ

$$P_s(\bar{N}, t)=\exp[-f\bar{N}]\cdot(f\bar{N}/t)^s\int_0^{t-(s-1)\tau} dt_1 \times\int_{t_1+\tau}^{t-(s-2)\tau} dt_2\cdots\int_{t_{s-1}+\tau}^{t} dt_s =\exp[-f\bar{N}]\frac{(f\bar{N})^s}{s!}\{1-(s-1)\tau/t\}^s$$

$$W(\bar{N}, t)=1-\exp[-f\bar{N}]\sum_{s=0}^{m+1}\frac{(f\bar{N})^s}{s!} \times\{1-(s-1)\tau/t\}^s;\quad m\tau\leqq t\leqq(m+1)\tau. \tag{2}$$

When $t\leqq\tau$ this equation becomes the same as (1) and constitutes the theoretical foundation of Talbots law. When $t>\tau$ we can derivate a formula that gives a fair approximation. The chance that the time between two absorptions is between T and $T+dT$ is $f\bar{N}/t\cdot\exp(-f\bar{N}/t\cdot T)dT$. The chance that light is percepted by these two absorptions is

$$\int_0^{\tau} f\bar{N}/t\cdot\exp(-f\bar{N}/t\cdot T)dT=1-\exp(-f\bar{N}/t\cdot\tau).$$

The chance for no light perception at all will be about $(\exp[-f\bar{N}/t\cdot\tau])^{f\bar{N}}\cong\exp[-(f\bar{N})^2\cdot\tau/t]$, as the average number of pairs of succeeding quanta is $f\bar{N}-1$ and $f\bar{N}\gg1$. So that

$$W(\bar{N}, t)=1-\exp[-(f\bar{N})^2\cdot\tau/t]. \tag{3}$$

For a chance of observation of 60 percent $\bar{N}_{60\text{ percent}}$ was calculated as a function of t/τ for $f=1$ with the aid of Eq. (2). Plotting $W(\bar{N}, t)$ against $\log\bar{N}_{60\text{ percent}}$, the shape of this function is independent of the value of f.

Obviously $\bar{N}_{60\text{ percent}}$ is constant for the case $k=1$. When $k=2$ it appears from Eqs. (2) and (3) that $\bar{N}_{60\text{ percent}}$ is proportional to $t^{\frac{1}{2}}$ for $t\gg\tau$. For the case that three quanta must be absorbed in the visual purple within a time τ, $\bar{N}_{60\text{ percent}}$ is proportional to $t^{\frac{1}{3}}$ for $t\gg\tau$, analoguous to $k=3$ under *C*. (See Fig. 1.)

Corresponding differences for probability curves representing form (1) are much smaller. The experimental results of van der Velden concerning $\bar{N}_{60\text{ percent}}$ for various values of t were satisfactory rendered by the 2-quanta case. In view of the great differences in the slope with t of $\bar{N}_{60\text{ percent}}$ for the cases $k=1$, 2, 3 it is impossible to conclude to one or three quanta. τ proved to be about 0.02 sec. The visual angle was about 4 minutes.

C. Determination of the Dependence of the Average Number of Quanta for a Fixed Chance of Observation on the Visual Angle of the Flash for a Short Time of Observation

In the case $k=2$ it is obvious that one has also to expect a dependence of $\bar{N}_{60\text{ percent}}$ on the visual angle. The condition caused by the absorption of a quantum will have an extension corresponding to the sensitive unit D of the retina. If the second quantum is absorbed outside the sensitive unit, activated by the first quantum no light is perceived. As long as the image of the light spot is small compared with the sensitive unit there will only be a slight dependence of $\bar{N}_{60\text{ percent}}$ on the visual angle d of the flash. Let $W(\bar{N}, d)$ be the chance of observation of a flash with visual angle d. When $d\ll D$

$$W(\bar{N}, d)=1-\exp(-f\bar{N})(1+f\bar{N}) -(f\bar{N})^2\cdot\exp(-f\bar{N})\cdot(\tfrac{2}{6}\cdot d/D-\tfrac{3}{6}d^2/D^2). \tag{4}$$

When $d\to0$ this equation merges into (1).

The chance for the flash falling completely within one sensitive unit and the chances for the falling of a definite fraction within one unit and a definite fraction within an adjoining unit are multiplied by the corresponding chances for a light perception. These results are added, and in this way (4) is obtained. It was assumed that the sensitive units as well as the light spot were square. For circular units the result will be only slightly different.

The course of $\bar{N}_{60\text{ percent}}$ with d according to (4) is only of slight importance as the deficiencies of the optical system of the eye obliterate its influence when $d<2$ minutes.

When $d \gg D$,

$$W(\bar{N}, d) = 1 - [\exp(-f\bar{N}\cdot D^2/d^2)\cdot(1+f\bar{N}\cdot D^2/d^2)]^{d^2/D^2} \quad (5)$$

as $\exp(-f\bar{N}\cdot D^2/d^2)\cdot(1+f\bar{N}\cdot D^2/d^2)$ is the chance for no light perception of a sensitive unit. The number of sensitive units is d^2/D^2. $\bar{N}_{60\text{ percent}}$ was calculated as a function of d by means of (4) and (5). Plotting $W(\bar{N}, d)$ against $\log\bar{N}$ the shape of this function is again independent of the value of f. As appears from (5), $\bar{N}_{60\text{ percent}}$ is proportional to d for $k=2$ and $d \gg D$. For (5) can be reduced when $d \gg D$,

$$W(\bar{N}, d) = 1 - \left[\left(1 - f\bar{N}\cdot D^2/d^2 + \frac{(f\bar{N})^2}{2!}\cdot D^4/d^4 \cdots\right)(1+f\bar{N}\cdot D^2/d^2)\right]^{d^2/D^2}$$

as $f\bar{N}D^2/d^2 < 1$ so that

$$W(\bar{N}, d) \backsim 1 - \exp(-\tfrac{1}{2}(f\bar{N})^2\cdot D^2/d^2). \quad (6)$$

In exactly the same way it can be proved that $\bar{N}_{60\text{ percent}}$ is proportional to $d^{2(k-1)/k}$ for arbitrary values of k. Analogous $W(\bar{N}, t)$ is proportional to $t^{(k-1)/k}$ when $t \gg \tau$. For the case $k=1$, $\bar{N}_{60\text{ percent}}$ is independent of d, so that again great differences occur in the slope of $\bar{N}_{60\text{ percent}}$ with d for $d \gg D$, for $k=1, 2, 3$. The experimental results of van der Velden were satisfactorally rendered by $k=2$. D proved to be 10 minutes. The flash time was about 0.01 sec. (see Fig. 2).

Equations (4) and (5) constitute the theoretical foundation of the laws of Ricco and Piper.

2. SOME REMARKS OF THEORETICAL KIND ABOUT THE TWO-QUANTA EXPLANATION AND THE DISCUSSION OF EXPERIMENTS OF OTHER INVESTIGATORS

After the war we received the papers dealing with experiments partly of the same kind as ours by Hecht[3] and co-workers. Only a fraction of the quanta incident on the eye is absorbed by the visual purple so as to give an impression of light. About half of the amount is lost by reflection and absorption in the various parts of the eye. Moreover not all quanta reaching the retina will be absorbed so as to contribute to the light perception. Hecht determined this effective fraction reaching the retina with the aid of threshold values for various wave-lengths and concluded that as an upper limit 5–14 quanta must be absorbed in the visual purple. His experimental results point to 1 quantum as the lower limit. From this part of Hecht's work it can be derived, that the number of absorbed quanta necessary for the light perception is at least one and does not exceed 14.

For the exact determination of this number Hecht used only the method described under *A* in Section 1, applying a visual angle of 10 minutes and a time of about 0.004 seconds. From the shape of the resulting curve they concluded that the cooperation of at least 5, 6, or 7 quanta (dependent on the observer) were necessary for the perception of light which disagrees with the results of van der Velden according to which

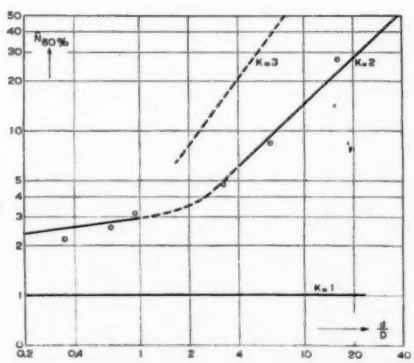

FIG. 2. The theoretical number of quanta, necessary for a chance of observation of 60 percent as a function of the visual angle d of the light spot. The theoretical values are given for the cases, that 1, 2 and 3 quanta are necessary for the perception of light. The experimental results for $t \leqq \tau$ from Fig. 7b reduced to $f=1$ are given.

independent of the observer two quanta are sufficient.

In our opinion there are several reasons for the two-quanta explanation being the most plausible, which reasons may here be briefly set forth.

Hecht and his co-workers used the same eye for the measurement of threshold values and for the observation of the red light for the fixation of the eye. From our experience however,[1,2] it was obvious that the red fixation light affects the observation of the light flashes, especially when the chance of observation is small. Consequently one obtains under these conditions a steeper probability curve than in the case that care is taken to observe the flash of light and the fixation light separately as was done in our experiments. Without this precaution the required number of quanta will obviously turn out to be higher.

For the exact determination of the number k method A described in Section 1 and used by Hecht is less suitable compared with the methods B and C. Besides the smaller usefulness caused by the smaller differences in the slope of the curves for the various cases $k=1, 2, 3, \cdots$ one can make objections of a theoretical nature against the method as soon as k is proved to be not equal to one.

It was assumed for all formulae of Section 1 when two quanta are absorbed in the visual purple the chance for a light perception is independent of the time ΔT between the two absorptions when $\Delta T \leqq \tau$ and is zero when $\Delta T > \tau$. It is very probable that this is only approximately good and a more continuous decrease of this chance $W(\Delta T)$ for ΔT values near τ will occur for increasing ΔT.

It is easy to understand that the slopes of the curves of method A will be steeper for a definite value of k by this matter. The chance for a light perception for a continuous course of $W(\Delta T)$ increases faster with increasing $\bar{N}$ of the flash as the chance for a light perception increases when the average time between the absorptions of the quanta decreases.

It was also assumed for the formulae of Section 1 that when two quanta are absorbed in the visual purple the chance for a light perception is independent of the distance between the two absorptions when the two quanta fall within a sensitive unit of diameter D. According to (4) the slope of the curve representing the chance function of method A depends on the visual angle of the flashes and is always steeper than Eq. (1) predicts. Moreover it is very probable that the chance for a light perception is not independent of the distance Δd between two absorptions of the two quanta fallen within a sensitive unit D, analogous to the facts for the dependence of the time between the two absorptions. By this fact the slope of the curve of method A is again steeper. (See Fig. 3.)

So the number k obtained from the experiments described in method A with the aid of formula (1) is the upper limit of the real number k. Only when the slope of the curve agrees with the case $k=1$ it is evident that these difficulties do not exist and no further informations are needed.[4]

It is not possible to conclude to the actual existence of definitely bounded sensitive units in the retina from the dependence of the slope of the chance function on the visual angle according to Eq. (4). It might be possible that the two quanta have to be absorbed within a distance D of each other instead of the necessity of two absorptions within an area of diameter D in the retina. Only the continuous decrease of the chance for a light perception with increasing distance of the absorptions of the two quanta $W(\Delta d)$ can cause the dependence of the slope of the chance functions of Fig. 3 on the visual angle. The formulae (4), (5), and (6) are derivated for the case that the sensitive units did exist in the retina. When for the light perception two quanta must be absorbed within a distance D of each other a formula similar to (2) is obtained. As the problem is now

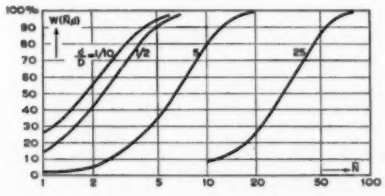

FIG. 3. Chance of observation as a function of $\bar{N}$ according to the two-quanta hypothesis for different values of d/D and $f=1$.

[4] E. C. Wassink and M. A. Bouman, Enzymologica 12 (1947).

two-dimensional d^2 corresponds to t, and the slope of the threshold values $\bar{N}_{60\text{ percent}}$ with d when $d \gg D$ is not influenced by this matter. For that reason we cannot conclude to the existence of the sensitive units from the experiments of the methods A, B, or C.

The number k obtained from the curve of method A will be closest the real number k when the visual angle of the light spot and the flash time are as small as possible. The influence of the facts just mentioned is in these conditions the smallest.

Hecht[3] applied a visual angle of 10 minutes and a flash time of about 0.004 seconds. As this visual angle is not small compared with D, the number of quanta necessary for the light perception was bound to come out too high. The influence of the visual angle on the slope of the curve of method A is also demonstrated by the experiments of C. Peyrou and H. Piatier,[5] who concluded to the value 2–5 for k. The deviations in their data from the theoretical two-quanta slope of Eq. (1) are still greater than ours, even for the smallest visual angle of 2 minutes. The deviations in our data (see also Section 3) almost agrees with Eq. (4). Unfortunately the authors did not give information about the manner of fixation of the eye and the spot of the retina on which the flashes were received. In Hecht's work the flash was seen 20 degrees nasal from the fovea, in ours 7 degrees.

It is very probable that the condition of the nerve system and the spot of the retina used in the observation of the flashes influence the chance for a light perception when two quanta are absorbed as function of the time and distance of these two absorptions: $W(\Delta T, \Delta d)$. These facts can cause a variability in the resulting curves of method A.

The two-quanta explanation was not only corroborated by the probability curves for the observation of light flashes of method A. Indeed, methods B and C are much more valuable for an exact determination of k, and for the test person of the present investigation, the results for $t<3\tau$ and $d<2D$ can only be explained by the two-quanta case, not only when method A was applied but also when methods B and C were applied.

It can be shown that for a continuous course of

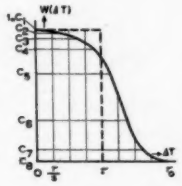

FIG. 4. Chance for a light perception when two quanta are effectively absorbed as a function of the time ΔT between the absorptions.

the chance for a light perception $W(\Delta T, \Delta d)$ the slope of the threshold curves $\bar{N}_{60\text{ percent}}$ of the methods B and C as far as $d \gg D$ and $t \gg \tau$ are still quite similar to the corresponding slopes of Eqs. (2), (3), (5), and (6), which were derivated for the chance function $W(\Delta T, \Delta d)$ represented by the dotted line of Fig. 4. We give the derivation for the time relation. For the visual-angle relation the facts are analogous. For an arbitrary course with time represented by the full line of Fig. 4, we divide the segment τ_0 in a large number of equal parts τ_0/s with corresponding average value of $W(\Delta T, \Delta d)$: c_k. The chance that the time between two succeeding absorptions is between ΔT and $\Delta T + d(\Delta T)$ is $f\bar{N}/t \cdot \exp(-f\bar{N}/t \cdot \Delta T) \cdot d(\Delta T)$. The chance for a light perception by these two absorptions is

$$\int_0^{\tau_0/s} f\bar{N}/t \cdot \exp(-f\bar{N}/t \cdot \Delta T) \cdot c_1 d(\Delta T) + \int_{\tau_0/s}^{2\tau_0/s} f\bar{N}/t \cdot \exp(-f\bar{N}/t \cdot \Delta T) \cdot c_2 d(\Delta T) + \cdots + \int_{(s-1)\tau_0/s}^{\tau_0} f\bar{N}/t \cdot \exp(-f\bar{N}/t \cdot \Delta T) \cdot c_s \cdot d(\Delta T) = 1 - (c_1 - c_2)\exp(-\tau_0/s \cdot f\bar{N}/t) - (c_2 - c_3)\exp(-2\tau_0/s \cdot f\bar{N}/t) \cdots - c_s \exp(-f\bar{N}/t \cdot \tau_0).$$

As the average number of succeeding pairs of quanta is $f\bar{N} - 1$, the chance for no light perception at all is about

$$[(c_1 - c_2)(1 - \tau_0/s \cdot f\bar{N}/t) + (c_2 - c_3)(1 - 2\tau_0/s \cdot f\bar{N}/t) + \cdots + c_s(1 - \tau_0 f\bar{N}/t)]^{f\bar{N}} \sim \left[c_1 - \frac{\sum c_k}{s}\tau_0 f\bar{N}/t\right]^{f\bar{N}}$$

[5] C. Peyrou and H. Piatier, C. F. de Soc. de l'Academie des Sciences F 113, 589 (1946).

so that

$$W(\bar{N}, t) = 1 - c_1^{f\bar{N}} \cdot \exp[-(\sum c_k/sc_1) \cdot \tau_0 (f\bar{N})^2/t]$$

when $\tau_0 \ll f\bar{N}/t$, analogous to form 3 as c_1 can always be chosen equal to one therefore, only the methods B and C are suitable for the exact determination of the number k.

There is still one more reason for the probability that the conclusion of Hecht for the necessity of 5, 6, or 7 quanta is too high. For one of the observers[1,2] was found $f = 28$ percent at $\lambda = 5050$A, on the suppositions that the two-quanta hypothesis is correct and that the course of the threshold values with wave-length is always in agreement with the international scotopic luminosity curve. Be that as it may, for $d = 4$ minutes, $t = 0.01$ seconds, and $\lambda = 5050$A, for this observer $\bar{N}_{60\ \text{percent}}$ never did exceed 14, as will be clear from Section 3. Since about half the amount of light incident on the eye is lost by reflection and absorption in its various parts, the upper limit of the average number of quanta reaching the retina is 7. Only a part of these quanta will be absorbed in the visual purple so as to give a light perception. If 5, 6, or 7 quanta would be necessary this fraction would have to be improbably large, as Wald[6] found this fraction to be of the order of magnitude of 20 percent.

In our experiments the absolute value of the energy is of no importance for the determination of the number k. Only the value of f of the test person is based on the absolute values. As we determined these absolute level in an indirect way, the error may be large but will certainly not exceed 25 percent (Section 3). For most of the test persons f has under optimal conditions a value of 7 to 10 percent, so that $\bar{N}_{60\ \text{percent}}$ when $t \ll \tau$ and $d \ll D$ is about 25 quanta. As far as they proved to be reliable the corresponding numbers in the work of other investigators are for the greater part almost twice as large. Hecht[3] examined all the published measurements on the energy required for vision. The data of Chariton and Lea yield 18 quanta, von Kries and Eyster give 34–68, Barnes and Czerny's work results in 40–87, Hecht found 58–148, and Peyrou and Piatier yield 70–90 quanta. The differences may be partly due to the variability of the absorbing and reflecting properties of the human eye. Moreover, the experimental arrangements of the several authors were not the same. In this connection the location of the illuminated area of the retina, the way of fixating the eyes and the general comfort of the observer are of special importance. Another paper was published during the war by Pirenne[7] in continuation of the work of Hecht. These experiments were performed in order to find out if there is any cooperation between the two eyes for the perception of light. His experiments for the left and right eye are not sufficiently complete to determine the number of quanta necessary for the perception of light. Besides this, Pirenne has used white light, so that his measurements are not strictly comparable to ours. He found that there was no cooperation between the two eyes. This is in agreement with our result that the two-quanta have to be absorbed within a distance $D(10')$ of each other.

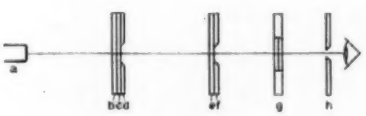

FIG. 5.

De Vries' estimation[8] of the number of quanta to be absorbed at the threshold of vision is not very convincing, since the results of different authors are used. The relation derived by him between contrast sensitivity and brightness for low values of the intensity using the one-quantum hypothesis, is not in accordance with the experiments of Steinhardt.[9]

Steinhardt found for low brightness values proportionality between the contrast sensitivity and the brightness, not the square root of the brightness, whereas de Vries derived proportionality with the square root of the brightness. An explanation on the basis of two quanta, will be published soon.

3. THRESHOLD-MEASUREMENTS WITH $t > \tau$ AND $d > D$

In this section we will present some further experiments relating to the behaviour of the thresh-

[6] G. Wald, J. Gen. Physiol **21**, 795 (1938).

[7] M. H. Pirenne, Nature **152**, 698 (1943).
[8] Hl. de Vries, Physica **10**, 553 (1943).
[9] J. Steinhardt, J. Gen. Physiol. **20**, 185 (1936).

old values when the flash time $t > \tau$ together with the visual angle $d > D$.

The experimental arrangement was the same as in the previous paper.[1,2] A tungsten ribbon filament lamp a was used as a light source. In order to get a sufficiently low brightness two opal glasses c and e were placed into the beam. The size of the light spot was controlled by interchangeable diaphragms f. A Schott Vg2 filter b with maximum transmission at 5300A° determined the color of the beam (see Fig. 5). The head of the observer was fixed by means of a mouthpiece mounting. The place of the retina where the flash was received was about 7° nasal from the fovea of the dark-adapted right eye. This fixation was realized with the aid of a red fixation light observed by the left eye. The aperture of the pupil of the right eye was reduced to a constant size by the intercalation of a diaphragm h.

The duration of the flash could be varied by means of a compur shutter g, which was regularly checked with the aid of an alternating light source, photo-cell, amplifier, and oscillograph. The properties of the opal glasses were studied with a photo-cell. The transmission curve of the filters had been determined with the photoelectric a.c. amplifier with a.c. galvanometer described by Milatz[10] and Milatz and Bloembergen.[11] The tungsten ribbon filament lamp is calibrated according to the principles developed and used by Ornstein and co-workers[12] in our institute.

The number of quanta per second incident on the eye for every wave-length is further provided by the strength of the current of the lamp and the geometry of the experimental arrangement. The intensity of the flash was varied by changing the distance and the current of the lamp. The error in the number of quanta incident on the eye may be about 25 percent but is for all strengths of the current and all distances exactly the same and cannot influence the slope of the curves on which the two-quanta explanation was based. Only the values of f depends on this error. As we did not use monochromatic light the quanta intensity of every wave-length was given a special weight in accordance with the international scotopic luminosity curve. This curve was normalized so as to make its optimal value equal to one; so that f in our experiments refers to the wave-length of the optimal value of the luminosity curve. This wave-length coincides with the optimal power of absorption of the visual purple.

The actual course of the threshold value with wave-length may differ from the international curve. The way in which the energies mentioned change with the current is practically not influenced by the differences, which may occur in

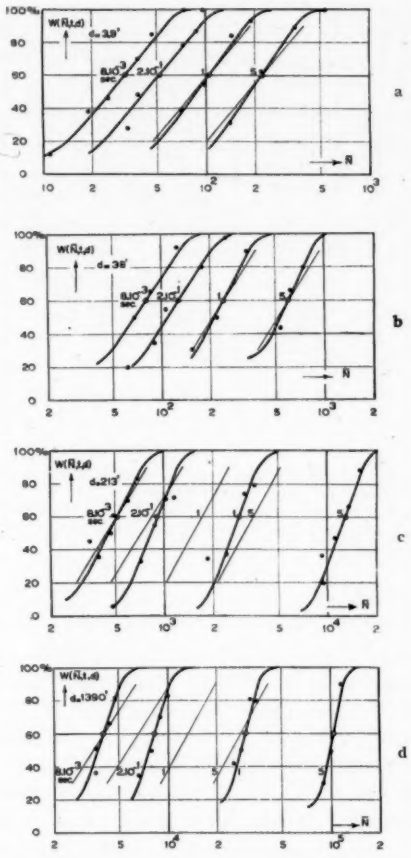

FIG. 6abcd. Theoretical and experimental (thick) curve for $W(\bar{N}, t, d)$, the chance of observation as a function of $\bar{N}$ for different combinations of d and t. The theoretical lines are adapted to the experimental ones by making them coincide at $W(\bar{N}, t, d)_{60\ \text{percent}}$ and $t = 8.10^{-3}$ sec.

[10] J. M. W. Milatz and L. S. Ornstein, Physica **2**, 355 (1935).
[11] J. M. W. Milatz and N. Bloembergen, Physica **11**, 449 (1946).
[12] L. S. Ornstein, W. J. H. Moll, H. C. Burger, Objektive Spektralphometrie Sammlung Göschen.; L. S. Ornstein, J. Opt. Soc. Am. **20**, 573 (1930); L. S. Ornstein, Mlle. J. G. Eymers, D. Vermeulen, Revue d'Optique 385 (1933).

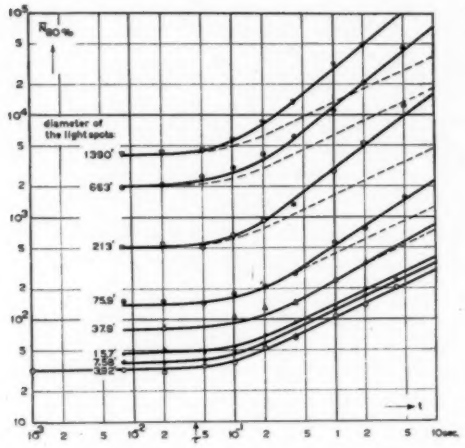

FIG. 7a. The number of quanta, necessary for a chance of observation of 60 percent as a function of t, the duration of the flash, again both theoretical (dotted curve) as well as experimentally (full-drawn curve), for various values of the diameter of the light spot.

the threshold curve. These only influence the value of f. The differences between the real threshold curve and the international curve increase with increasing absorption capacity (Hecht[3]). When our experiments result in f-values of about 10 percent or smaller, the influence is negligible. Taking into account the possible error in the absolute level of the energy and in the course of the threshold curve with wave-length, for the one observer for whom $f=28$ percent was found $\bar{N}_{60\text{ percent}}$ under optimal conditions will not exceed 14 quanta. This number is obtained by assuming the eye to be equally sensitive for all wave-lengths, so that it is the real number of quanta incident on the eye.

For many combinations of values of t and d the chance for observation, $W(\bar{N}, t, d)$, of the flash was determined as a function of $\bar{N}$. The values of $\bar{N}$ for an observation chance of 60 percent was determined from these measurements. Using the formulae (2) and (5) we obtain the theoretical expression for $W(\bar{N}, t, d)$ for arbitrary values of t and for $d \gg D$

$$W(\bar{N}, t, d) = 1 - \left\{ \exp(-f\bar{N}D^2/d^2) \times \sum_{s=0}^{m+1} \frac{f\bar{N}D^2/d^2}{s!} [1-(s-1)\tau/t]^s \right\}^{d^2/D^2} \qquad (7)$$

in which $m\tau < t \leqq (m+1)\tau$.

The theoretical values of $W(\bar{N}, t, d)$ and $\bar{N}_{60\text{ percent}}$ (the number of quanta necessary for a chance of observation of 60 percent) were derived from Eqs. (2), (4), (5) and (7).

The theoretical and experimental values of $W(\bar{N}, t, d)$ and $\bar{N}$ are given in Figs. 6a–d and the results for $\bar{N}_{60\text{ percent}}$ in Figs. 7a and 7b.

4. DISCUSSION OF THE EXPERIMENTAL RESULTS

On repeating the experiments of van der Velden[1,2] with another test person, we found again that the experimental $W(\bar{N})$ curve, in the case that $t<\tau$ and $d<D$, was in accordance with the two-quanta explanation.

Both the thresholds as a function of d, when $t<\tau$ and the thresholds as a function of t when $d \ll D$ are also fully covered by the two-quanta case, as can be seen from Figs. 7a and 7b. For the test person (M.A.B.) used in these experiments we found $D=12'$, $\tau=0.035$ sec. and $f=8$ percent. The value of τ is a little higher than for H. v.d. V. in the previous experiments (τ about 0.02 sec.).

When $t>3\tau$ together with $d>3D$ considerable deviations from the theoretical values are found in the probability and threshold curves. The threshold values show a much steeper dependence upon t and d than the two-quanta assumption requires. Such a dependence (proportional with d^2 respectively t) was already known from earlier publications. As previous authors do not take into account the statistical character of vision, so that the threshold value could not be defined exactly, their measurements are unsuitable for quantitative comparison with ours.

In this paragraph we shall show what causes cannot be responsible for the deviations, mentioned below under (a)–(f), from the theoretical curves and how we have consequently been forced to the explanation given at the end of this section.

(a). Variation of f, D or the number of rods per mm² in the region of the retina involved in the experiments. According to Østerberg[13] the number of rods per mm² is a function of the place on the retina. It is obvious, however, that this cannot explain the influence of t on the slope of the threshold curves. It could only lead to the introduction of a correction to the slope of the threshold curve as a function of d, independent of the value of t. The theoretical curves in Fig. 7b are corrected for the variation mentioned, the

[13] Østerberg, Acta Opthalm **13** (Supplement), 1935.

correction turned out to be small. When D and f are independent of the place, $\bar{N}_{60\text{ percent}}$ must be exactly proportional to d in Fig. 7b.

(b). Variation of τ. If variations of τ in the region of the retina, involved in the experiments, would furnish an explanation for the deviations found, the experimental curve of the threshold as a function of t would lie between the theoretical curves corresponding to the two extreme values of τ found in the region involved. To explain the deviations found, these two extremes should differ at least by a factor 50. Measurements of τ, on different places of the retina, gave at most variations of 50 percent. For that reason these variations cannot explain our experimental results.

(c). When the number of quanta necessary for the light perception is not constant and varies from time to time and there exists a frequency distribution for the various cases of k, the slope of the threshold curves $\bar{N}_{60\text{ percent}}$ as a function of d when $t \gg \tau$ and of t when $d \gg D$ will become in agreement with the smallest possible number of k, when t respectively d is large enough. The change in slope of the curves mentioned for increasing t respectively d is just in the opposite direction and changes from the slope corresponding to the case $k=2$ to almost $k=3$.

(d). The slope of the threshold curves when $t>0.2$ sec. and $d>20'$ seems to be in better agreement with a three-quanta hypothesis. However, if besides the two-quanta condition an extra condition necessary for the perception of light is introduced, as, for example, the absorption of a third quantum within τ' sec. after the absorption of the first two quanta, it is obvious that this new condition would influence all the threshold curves as a function of t, d being constant and smaller than $2D$, and the threshold curve as a function of d when t is constant and smaller than 3τ. However, these threshold curves can only be explained by the simple two-quanta case.

(e). The hypothesis of de Vries,[8] that the retina grey should be of importance in the region of low brightness is not probable. We found that when flashes light up for a test person with an average number of quanta unknown to him, he never has an impression of light when the number of quanta is zero. It seems, therefore, probable that the retina grey does not give the same impression as a normal flash of light.

(f). The variation of the concentration of the visual purple with time cannot play an important role, because, taking for example the case that $d=100'$ and $t=5$ sec., only 170 quanta are absorbed by the visual purple. Assuming an average of 100 rods in a sensitive unit in this area with $d=100'$ the total number of rods is about 6500, so that only 2.6 percent of the rods have absorbed a quantum. During the flash this percentage is still smaller and decreases with increasing diameter and duration of the flash. In view of the fact that the rods cannot influence each other except for their connection by the nerve system, it is not conceivable, that a change in the concentration of the visual purple in one rod, causes such a change in another rod. The deviation mentioned needs more than 60 percent for this percentage.

After having discussed these various phenomena, we shall now give an explanation which seems to us to be the most probable one. As was already found[1,2] and is corroborated in this paper, two quanta have to be absorbed by the visual purple within τ sec. and within an area corresponding to a visual angle D, in order to give a light impression when the durations of the flash or the visual angles are small. In a sensitive unit with a visual angle D there are about 100 rods. As the chance is negligible, that both quanta are absorbed in the same rod, it is clear that after the absorption of one quantum there must be already a certain change: an impulse will be transmitted by the nerve connection of a rod when one quantum is absorbed in it. Now one gets an impression of light when a second quantum is absorbed within a visual angle D and within a time τ. Since the only way for interaction between two rods is their nerve connection, after the absorption of one quantum there will be some nerve change extending over a certain area of the retina. The nerve change caused by the absorbed quanta apparently gives rise after some time to such a condition of the apparatus for vision that the impression of light after the absorption of a pair of quanta is impeded. This state has developed after a time T (about 0.1 sec.) and extends over an area O (about 20'). It is not im-

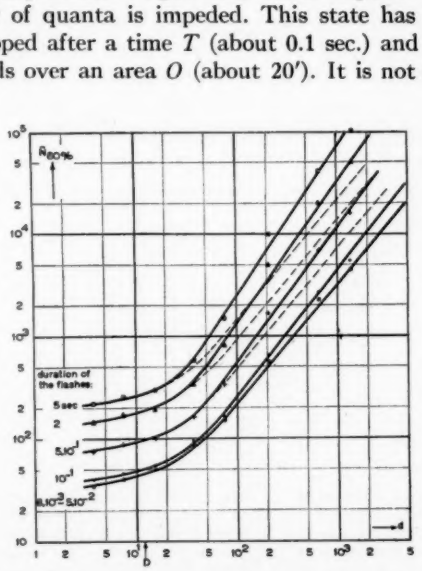

FIG. 7b. The number of quanta, necessary for a chance of observation of 60 percent as a function of the visual angle d of the light spot, both theoretical (dotted curve) and experimental (full-drawn curve), for various durations of the flash.

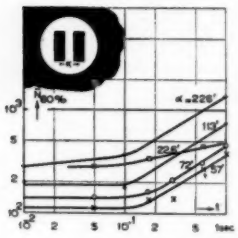

FIG. 8. The number of quanta from the field between the two black stripes as a function of t, the time of observation, for different values of the visual angle α, for a chance of recognition of 60 percent.

possible that this phenomenon is related to the α-adaption described by Schouten and Ornstein.[14]

The deviations of the threshold curves from the two-quanta theory are in accordance with this phenomenon. A further discussion will be given in another paper in connection with the problem of brightness impression and contrast sensitivity.

5. VISUAL ACUITY

One method, very often used for determinations of the visual acuity, consists of the recognition of two separate black stripes as given in Fig. 8. It is clear that visual acuity is closely related to threshold values. The two stripes will be recognized as separately when the field between the two black stripes is seen. Obviously we can define only a certain chance of recognition just as in the case of threshold values. Moreover, it is evident that this chance of separate recognition will show a similar dependence on the visual

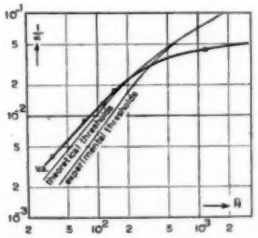

FIG. 9. The reciprocal of the visual angle α (visual acuity) as a function of $\bar{N}$, the brightness, expressed in the number of quanta per second and per square area (laying on the retina) with a visual angle of 10^{-2} radian. The theoretical curves are given, using the threshold values from the two-quanta theory and the threshold values found experimentally.

[14] J. F. Schouten and L. S. Ornstein, J. Opt. Soc. Am. **29**, 168 (1939).

angle and the time of observation as was established for the measurements of the threshold of perception. Again we take the chance of 60 percent as a criterion.

As soon as the visual angle of the details is of the same order of magnitude as the sensitive unit, deviations will occur from threshold curves of the types shown in Figs. 7a and 7b. These deviations are due to the fact that within a sensitive unit separate light sensations cannot be observed at a low brightness. With the same arrangement as was used for threshold measurement, we determined the relation between the chance of recognition, and the visual angle of the stripes and the duration of the flash. These measurements were performed for different values of the brightness. The spot of the retina during these experiments was again 7° nasal from the fovea.

The number of quanta from the field between the two black stripes is given in Fig. 8 as a function of the time of observation for different values of the visual angle covered by the field between the stripes. Comparisons were made for a chance of recognition of 60 percent. The curve for 22.9′ shows the existence of the sensitive unit. From these experimental curves, one can derive the relation between the reciprocal of the visual angle (the visual acuity) and the brightness for a fixed duration of the flash, for example, 1 second, as is shown in Fig. 9.

The analogous function was derived from the threshold measurements and from the theoretical threshold values from Fig. 7a and 7b, and is also given in Fig. 9.

From Fig. 9 we can conclude that the visual acuity is in better accordance with the thresholds derived from the two-quanta theory than with the thresholds found experimentally. The reason for this seems to be the influence of the illuminated peripherous part of the test figure.

From other experiments to be discussed in a subsequent paper we must draw the conclusion that when two quanta have cooperated somewhere to give an impression of light, the absorption of only one quantum in the neighborhood is sufficient to give rise to an extra impression of light. Of course this phenomenon counterbalances to a certain extent the influence of the phenomenon, described at the end of Section 4. As could be expected from these phenomena a better

agreement with the experimental thresholds was found when for measurements of visual acuity a figure was used of three illuminated stripes completely isolated from each other by dark areas.

Of course, the primary effect of an absorbed quantum in the visual purple of a rod is of a photochemical nature. The chance of observation of a light spot as well as the chance of recognition of a figure (so long as the visual angle is large compared with the sensitive unit) are described, both under various circumstances, by the statistical fluctuations in place and time of the absorptions. The two-quanta theory explains the facts of this behavior as satisfactorily as can be expected for an object like the human eye—both for rod and cone vision. In one of the subsequent papers we will give the data obtained for cone vision. A photochemical theory about threshold values and visual acuity as given by Hecht[15] and very recently by Jahn[16] can by no means give a complete account of the facts as far as the dark-adapted eye is concerned. For, in this theory the statistical nature of vision and the properties of the nerve system have completely been neglected.

De Vries[8] has given a theory of the visual acuity, using the one-quantum hypothesis. His relation between visual acuity and brightness is not in accordance with the experimental curve. According to the one-quantum theory, the threshold values are not dependent on the visual angle. It would be expected in that case that the visual acuity could be expected to be proportional to the square root[17] of the brightness which is contradictory to the experimental results.

The visual acuity can only be explained by a one-quantum hypothesis if the retina grey should have an essential influence. This however is not probable as was discussed in Section 3.

Finally, we may still remark that the experimental results of Graham and Cook[18] about visual acuity, and of Graham and Margaria[19] and of Wald[20] about visual threshold are in agreement with the results here presented in so far as their measurements, which neglected the consequences arising from the quantum nature of light, are comparable with ours.

Our sincere thanks are due to Professor J. M. W. Milatz for his kind and continual interest in this work and to the editor J. R. Harrison for valuable suggestions regarding the paper.

Note added in proof: We very recently received the paper of M. H. Pirenne,[21] and will discuss his work in a subsequent paper.

[15] S. Hecht, Physiol. Rev. **17**, 239 (1937).
[16] T. L. Jahn, J. Opt. Soc. Am. **36**, 83 (1945).
[17] Erroneously in his original paper proportionality between brightness and visual acuity is derived from the one-quantum hypothesis.
[18] C. H. Graham and C. Cook, Am. J. Psychol. **49**, 654 (1937).
[19] C. H. Graham and R. Margaria, Am. J. Physiol. **113**, 299 (1935).
[20] S. Wald, J. Gen. Physiol. **21**, 269 (1938).
[21] M. H. Pirenne, Proc. Camb. Phil. Soc. **42**, 78 (1946).

Emission Spectra of Some Zinc Sulfide and Zinc-Cadmium Sulfide Phosphors

W. H. Byler
United States Radium Corporation, New York, New York
(Received July 8, 1947)

Emission spectra of thirty zinc sulfide and zinc-cadmium sulfide phosphors with varied cadmium content and different activators are shown. The spectra were obtained with a grating spectrograph using photographic recording, and the curves are plotted in terms of relative energy. In spite of wide variation in cadmium content and use of different activators, peaks and plateaus appear repeatedly at certain positions. Analysis indicates that the broad emission bands are actually summations of contributions from a number of individual bands whose peak positions are invariant. Since these bands are found in the samples containing no added activator and since all of the activated samples show them in certain wave-length ranges, it is concluded that they are characteristic of the base material rather than the activator. Bands peaked at the following approximate positions were found: 430, 475, 520, 555, 575, 600, and 625 millimicrons. It is suggested that these different bands are related to differences in types of atomic groupings in the base material. The activator serves to influence the relative heights of the bands and may accomplish this through a differential influence on the ratio of radiative to non-radiative transitions in the different centers

IT seems to be generally believed that the emission spectra of the zinc sulfide and zinc-cadmium sulfide phosphors are characterized by single broad bands whenever a single activator at optimum concentration or no activator is used.[1, 2] Also, it is thought that each activator used in zinc sulfide introduces its own characteristic emission band and that the peak position of this band is shifted gradually toward longer wave-lengths as more cadmium is introduced. Contrary to these views, results of many measurements in this laboratory seem to indicate that the broad emission band is in every case a summation of contributions from a number of individual bands whose peak positions are invariant and are characteristic of the base material rather than the activator.

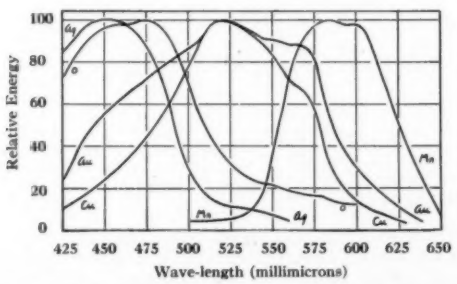

FIG. 1. Emission spectra of zinc sulfide phosphors with following added activators: none, silver, gold, copper, and manganese.

For the purpose of this presentation, a group of thirty phosphors was prepared. The cadmium sulfide content was varied from 0 to 50 percent and each sample contained either no added activator or a single activator, silver, gold, copper, or manganese. The activator content was 0.01 percent for silver, gold, and copper, and 0.5 percent for manganese. Sodium chloride served as flux in each case and the roasting temperature was varied from 1085°C to 1130°C in order to yield approximately uniform average crystal size.

The emission spectra, from excitation by P-7 type Sylvania Blacklite lamp with Corning filter No. 587, were recorded with a Cenco grating spectrograph with dispersion of 16A/mm. The method for calibrating the film has been described elsewhere[3] and involves use of two phosphors with known emission spectra as standards. The spectra of these standards were originally determined with a visual spectrophotometer using a calibrated lamp. For this work, some errors from the visual work were corrected by recording the two phosphor spectra and the spectrum of an incandescent lamp of known color temperature on the same film.

[1] H. W. Leverenz, Proc. I.R.E. **32**, 256 (1944).
[2] F. E. Williams, J. Opt. Soc. Am. **37**, 302 (1947).
[3] W. H. Byler and C. C. Carroll, J. Opt. Soc. Am. **35**, 258 (1945).

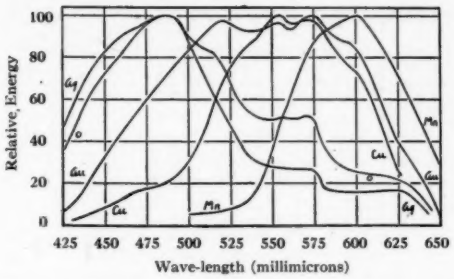

FIG. 2. Emission spectra of zinc-cadmium sulfide phosphors, ZnS(90)Cds(10) with following added activators: none, silver, gold, copper, and manganese.

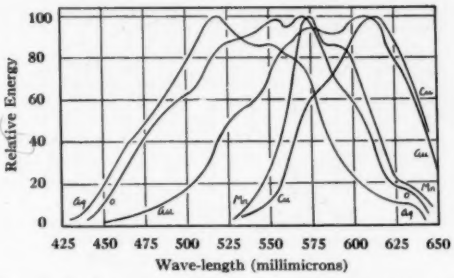

FIG. 4. Emission spectra of zinc-cadmium sulfide phosphors, ZnS(70)CdS(30) with following added activators: none, silver, gold, copper, and manganese.

Having once determined the relative energy curves for the phosphors, one can use these repeatedly for calibrating new film. Thus, each film used in this work was calibrated. It is considered advantageous to use standard phosphors for calibrating the film since this provides conditions similar to those employed for the unknowns and it avoids fluctuations to which a lamp standard may be subject.

The results are shown in Figs. 1–6. The striking feature of these curves is the repeated appearance of peaks and inflections at certain positions. The second significant observation is that all of these band positions are shown by the non-activated samples. Finally, each activator reveals all of the bands in the wave-length range where it is effective.

Conclusive identification of all of the suggested bands may require spectra of a larger number of samples. Also, the suggestion that the band positions are invariant does not at first glance appear to be supported by these curves. But it is obvious that addition of two broad bands whose peaks are sufficiently close together can produce a curve with a false peak position between those of the contributing bands. Assuming that the curves are summations of contributions from individual bands, we can proceed by trial and error to locate the bands. Some of the peak positions are repeated so frequently that we can identify some of the bands with assurance. Having a few bands well established, location of

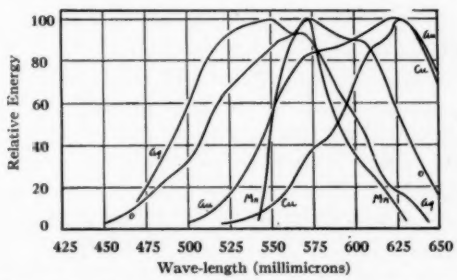

FIG. 5. Emission spectra of zinc-cadmium sulfide phosphors, ZnS(60)CdS(40) with following added activators: none, silver, gold, copper, and manganese.

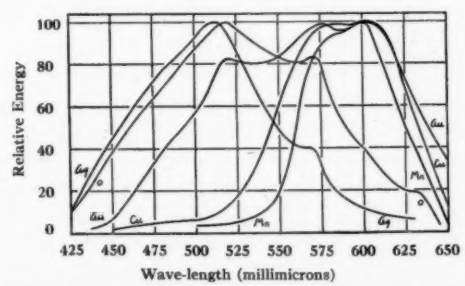

FIG. 3. Emission spectra of zinc-cadmium sulfide phosphors, ZnS(80)CdS(20) with following added activators: none, silver, gold, copper, and manganese.

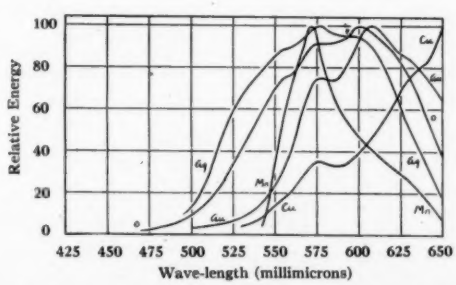

FIG. 6. Emission spectra of zinc-cadmium sulfide phosphors, ZnS(50)CdS(50) with following added activators: none, silver, gold, copper, and manganese.

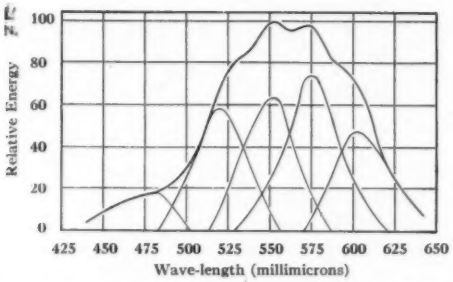

FIG. 7. Analysis of the emission spectrum of ZnS(90)-CdS(10)Cu. Addition of the individual bands yields the measured curve.

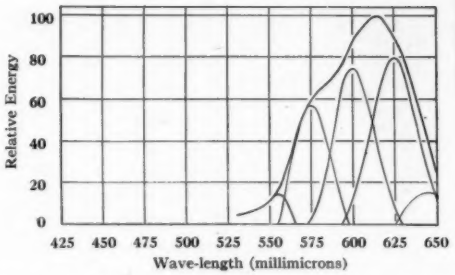

FIG. 8. Analysis of the emission spectrum of ZnS(70)-CdS(30)Cu. Addition of the individual bands yields the measured curve.

the others is simplified. Two such analyses are shown in Figs. 7 and 8. The end portions of the curves are too low lying for analysis, but the principal bands shown are either symmetrical or nearly so. Any departure from true symmetry may be due to errors in the original curve measurements. There is less material with which to work in the blue end of the range but similar analyses there point to bands peaked at about 430 and 475 millimicrons. Casual examination of the emission curves leads to expectation of bands at around 450 and 485 millimicrons, but analysis favors the other positions. These results, therefore, point to individual emission bands peaked at the following approximate positions: 430, 475, 520, 555, 575, 600, and 625 millimicrons. It is further indicated that there are other bands beyond the range under consideration.

These results raise important questions concerning the nature of the active centers and the role of the activator in these phosphors. The activators appear to exert only a secondary effect on the emission spectra in that they only influence the relative intensities of bands which are there anyway. Interstitial zinc is generally considered to serve as activator in the absence of other activators. Whether the emission bands are directly associated with the interstitial zinc atoms or whether zinc serves in the same capacity as the other activators is a question not answered by this work. However, since there are multiple bands and only one kind of interstitial atom in this case, we are led to conclude that there are different types of centers which differ in atomic groupings of the base material. This is supported by the fact that the relative heights of the independent bands are related to amount of heat treatment applied in preparation of the phosphor. Many examples are known[4] in which the only necessary condition for development of luminescence appears to be creation of lattice distortion.

The results presented in this paper show that a given activator favors the emission of some bands over others in a given base material of the zinc-cadmium sulfide class. Whether this is accomplished through favoring formation of certain center types or by some other mechanism is not revealed. An alternative mechanism would be a differential influence on the ratio of radiative to non-radiative transitions in the different centers. It is hoped that further investigation will throw light on this question.

[4] W. H. Byler, J. Am. Chem. Soc. 60, 1247 (1938).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 11 NOVEMBER, 1947

An Infra-Red Spectrophotometer for Industrial Research

E. J. MARTIN, A. W. FISCHER, B. MANDEL, R. E. NUSBAUM
Research Laboratories Division, General Motors Corporation, Detroit, Michigan

(Received July 1, 1947)

An infra-red spectrophotometer has been designed for the wide variety of problems which arise in an industrial research laboratory. The source of radiation is a Nernst glower, the dispersing unit a 10×15 cm sodium chloride prism, and the detector is a compensated thermopile. The output of the thermopile is amplified with a contact modulated amplifier. The installation is comparatively insensitive to mechanical vibration and fast temperature changes. The instrument is enclosed in a vacuum tank to prevent interference from absorption by extraneous gases. Driving motors for prism rotation and slit adjustments are outside the vacuum. Rotary motion is transmitted through the walls of the vacuum tank by means of magnetic drives; this avoids the use of any form of packing gland. Ink-line records of the spectral range from 3 to 15 μ can be made automatically in 30 minutes. The instrument is operated semiautomatically for "spot work" in making quantitative analyses after spectral positions and proper calibrations have been made.

FOR many years the development of better fuels has been of mutual interest to the automotive and petroleum industries. Fuel development programs involve the study of hydrocarbons, and in order to make progress in such work, it is very desirable to have the best available methods of analyzing such compounds. In 1933 the possibility of using infra-red absorption spectroscopy as a method of analyzing motor fuels was investigated by Kettering and Sleator.[1] They examined the infra-red absorption spectra of twenty-five compounds of which seventeen were typical of the important components of gasoline. Their work covered the range from 1μ to 30μ.

Since that time many investigators have published information on the industrial application of infra-red spectroscopy. Some recent work carried on by the Research Division has been described in a paper on the analysis of several specific liquid mixtures.[2]

The introduction of analytical methods which involve the use of instruments such as the infra-red spectrophotometer requires that the instrument and its associated equipment operate satisfactorily in an industrial environment. When the need for the present instrument was recognized there were no commercially available units appropriate for the problem in question. This led to the design and construction of the spectrophotometer now to be described.

This instrument has sufficient dispersion, resolving power, and versatility to be used as a research tool and at the same time, incorporates convenience, speed, and other features which make it applicable to process control and other routine analysis work in an industrial laboratory. The present paper will be confined to a description of the optical system and mechanical construction; other papers will describe the recording system, electrical controls, and performance characteristics.

DESIGN OBJECTIVES

A view of the complete instrument is shown in Fig. 1. The objective of this design was to incorporate the following features:

1. An optical path having minimum absorption by extraneous gases.
2. Freedom from sensitivity to mechanical vibration.
3. Minimum effects from temperature changes and wearing of moving parts.
4. Convenience in alignment of optics and permanence of adjustment.
5. Convenient and rapid operation.
6. Centralization of controls and recording system.
7. Spectrum recorded automatically in a form both permanent and immediately available.

OPTICAL SYSTEM

The optical system is shown in Fig. 2. It is identical with one employed by the University

[1] Charles F. Kettering and W. W. Sleator, Physics **4**, 39 (1933).
[2] D. L. Fry, R. E. Nusbaum, and H. M. Randall, J. App. Phys. **17**, 150 (1946).

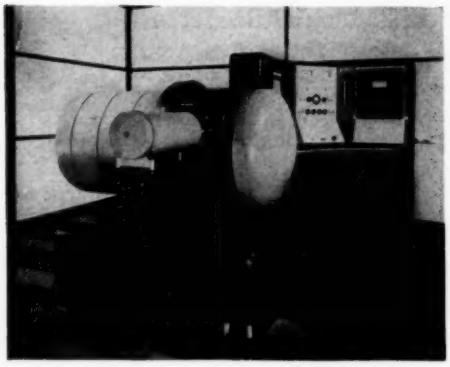

FIG. 1. Infra-red spectrophotometer. External form of the instrument was dictated primarily by functional requirements with emphasis on convenience and neatness.

of Michigan. A rock salt prism is used in the instrument; it was figured by the Gaertner Scientific Corporation. The mirrors were figured by the Perkin-Elmer Corporation. The slits are very similar to those previously described by Roemer and Oetjen.[3]

VACUUM TANK AND BASE

The spectrophotometer is housed in a vacuum tank as a means of reducing the absorption due to air and water vapor. The tank is a three-piece welded-steel cylinder having a center section with a short bell on one end and a longer bell on the other. The center section bolts to a heavy welded steel base, and supports in its interior a cantilever table upon which the optical system rests. The bells are easily removed by one man through the use of rails and carriages which may be temporarily installed and which permit the bells to be rolled back to allow free access to all parts of the spectrophotometer (see Figs 3 and 4). This general arrangement has the following desirable features:

1. The optical train is entirely independent of the bells and can be aligned and adjusted with the bells removed.
2. The necessary drives, wiring, vacuum-exhaust line, and entrance window can be permanently installed in the center section of the tank, and are in no way affected by the removal of the bells.
3. The shape of the bells allows maximum strength with minimum weight.
4. The circular tongue-and-groove seal is simple to machine and can be held to close tolerances.
5. The bells are easily rolled back and the bolt holes are always in line when the bells are replaced. Although the bells are quite heavy, they can be replaced without jarring the instrument.

In sealing the two bells, high-unit-pressure is obtained by using a narrow neoprene gasket and compressing it in a groove with a close fitting tongue to constrain the neoprene from flowing. This arrangement has been very effective and is used on all the vacuum seals required in the system. The vacuum tank can be pumped down to 100 microns, which is considered a satisfactory working pressure, in about five hours. Continu-

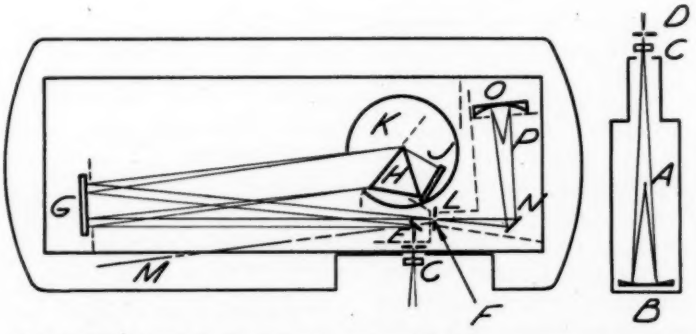

FIG. 2. Optical system. The system is conventional. The collimating mirror is a 10° off-axis paraboloid. The Littrow mirror moves with the prism. *A*, source; *B*, spherical mirror; *C*, absorption cell; *D*, entrance slit; *E*, plane mirror; *F*, focal point of 10° off-axis mirror; *G*, 10° off-axis parabolic mirror; *H*, prism; *J*, Littrow mirror; *K*, prism table; *L*, exit slit; *M*, axis of paraboloid; *N*, plane mirror; *O*, spherical mirror; *P*, thermopile.

[3] Hermann Roemer and Robert A. Oetjen, J. Opt. Soc. Am. **36**, 47 (1946).

ous pumping brings the pressure down to 18 microns. After the pressure has been reduced to this point, the pump can be turned off for about 40 hours before the pressure will rise above 100 microns. In other words, if the pump is operated every other night, the pressure in the tank remains within a satisfactory working range.

A welded steel base supports the spectrophotometer at a convenient working level and provides sufficient mass to keep the assembly from being top heavy. This base is in turn supported on two steel I-beams which rest on Lord rubber mounts. The room is provided with a raised floor which conceals these beams and mountings, and thus allows the supporting points to be widely spaced. Electrical leads, vacuum lines, and water pipes are brought out through the base and under this false floor to the control console or to the outside of the room, thus achieving the clean-cut appearance indicated in Fig. 1.

OPTICAL BED PLATE

Permanence of alignment of the optical parts depends very greatly upon the bed plate on which they are supported. This plate is made of 1″ invar, which has a very low coefficient of thermal expansion and is stress-relieved to minimize change of shape with time. The bed plate has three legs. One leg sits in a conically shaped hole and each of the other two rests on a large free ball, one on a plane and one in a groove, to give accurate positioning of the plate with respect to the source, and, at the same time, to allow contraction and expansion of the cantilever supporting table without introducing bending forces on the bed plate itself.

MIRROR MOUNTINGS

The form of the mirror mountings may be seen from Figs. 4 and 5. These mountings have the following features:

1. The mirror is supported on a short rod of such diameter that it will support the mirror, and at the same time can be deflected by means of four fine-pitch adjusting screws to give micrometer adjustment about horizontal or vertical axes for alignment of the mirror.
2. Any one of the three adjustments, horizontal, vertical, or focal, can be made independently of the other two.
3. The mountings are made of invar to prevent movement of mirrors due to temperature variation.

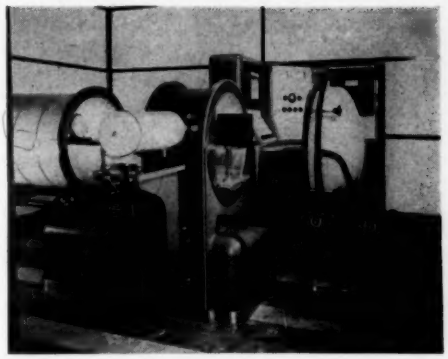

FIG. 3. Vacuum bells pulled back. Removable track and carriage system is used to withdraw the vacuum bells exposing the optical system.

4. All soldered joints are silver soldered in a hydrogen atmosphere furnace to avoid cracking or high local stresses.

PRISM TABLE

The prism table and drive require greater precision than any other mechanical part of the spectrophotometer, for, although the prism and Littrow mirror are rotated through an angle of 10 degrees, their position must be known at any point to within 1 second of arc. At the same time the axis of rotation of the table must remain strictly perpendicular to the optical axis. Two pair of radially matched duplex bearings (MRC 116-KRDT) of controlled quality were obtained from the Marlin Rockwell Corporation to serve as the bearings on the prism table. A large

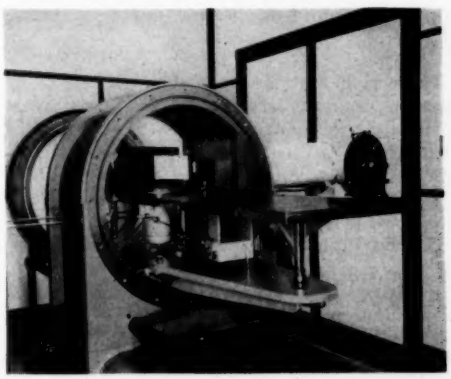

FIG. 4. Internal view of spectrophotometer. The invar optical base, supported on the cantilever member from the center section, is thermally isolated from the outer shell.

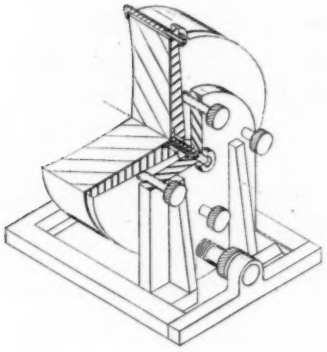

FIG. 5. Mirror mounting. The central rod supports the mirror in approximately its design position. Slight rotation around two perpendicular axes is effected with the screws by bending this rod.

diameter bearing, O.D.$=4\frac{15}{16}''$, having a large number of small diameter balls, was selected in order to reduce to a minimum the errors from bearing irregularities. The two pairs of bearings were mounted $7\frac{1}{4}''$ apart and assembled with a small preload to eliminate any tilting of the table. The high quality of the bearing was relied upon for the steadiness of the table. It was found that the table remained level for various angles of rotation to within the accuracy of measurement, which in this case was ± 5 seconds of arc. The outer shell of the bearing housing is visible in Fig. 6.

THERMAL INSULATION

The following precautions are taken to insulate the optical system from fluctuating room temperatures:

1. The vacuum tank is supported on the base by four hardwood blocks which hold the tank and base $\frac{1}{2}''$ apart.
2. The entire vacuum tank is insulated with $\frac{3}{4}''$ of felt. The felt is covered by a sheet metal envelope, providing a smooth surface for painting.
3. The flanges are covered with $\frac{1}{2}''$ of wood to provide insulation and to conceal the bolt heads.
4. The tank has a large thermal capacity which reduces the effects of rapidly fluctuating room temperatures.
5. The bed plate is supported on three legs which are thermally insulated from the cantilever table by micarta blocks.
6. The optical parts are inclosed in a highly polished radiation shield (not shown in the illustrations).

Although these precautions were taken to insulate the optical system, it was realized that there would be temperature changes and that some of these might be significant. In order to measure these changes, thermocouples (iron-constantan) were located at the following positions: the surface of each mirror including the one in the source box, each of two pair of slit jaws, each of two slit supports, the thermopile case, the source box, the metal shutter, and two positions on the optical bed plate.

These thermocouples can be connected one at a time to a potentiometer-type temperature indicator. In order to increase the accuracy of this instrument, a station is provided to check the cold junction compensation. This is done by comparing the temperature reading with that of a precision thermometer immersed in a thermos flask filled with oil. A thermocouple connected to the temperature indicator is also immersed in this flask. A cold junction correction amounting to the difference between the reading of the temperature indicator and that of the thermometer is applied to all readings.

A special arrangement has been made for the the prism temperature measurement. In order to obtain an average value, five junctions in series are placed on the surface of the prism. All the cold junctions except the one at the indicator are immersed in the thermos flask. The prism temperature then becomes $T_p=T_1+[(T_2-T_3)/5]$ where T_1 is the temperature of the flask as indicated by the thermometer, T_2 is the indicated temperature of the prism and T_3 is the indicated temperature of the flask.

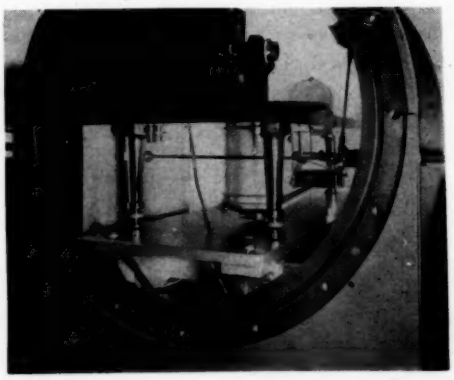

FIG. 6. Prism table and drive. The prism table is supported by preloaded ball bearings.

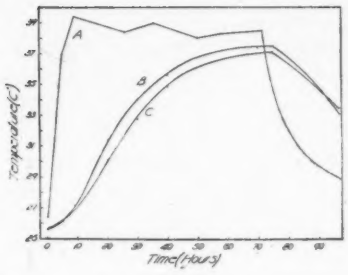

FIG. 7. Curves showing results of temperature study. *A*, average of the room temperature and parts in the source box; *B*, average temperature of parts inside the spectrophotometer; *C*, prism temperature.

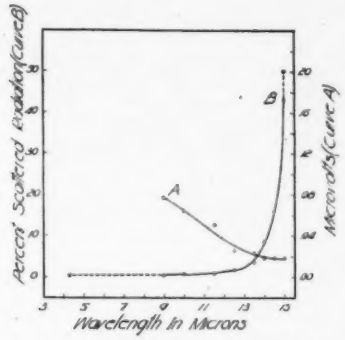

FIG. 8. Scattered radiation. *A*, thermopile response in microvolts due to scattered radiation (slits at 1.8 millimeters); *B*, percent scattered radiation.

Two series of observations on temperature changes have been made. The first was made with the Nernst glower turned off and no cooling water circulating through the source box. A room heater was turned on and operated for a period of seventy-two hours; a constant room temperature was reached in about nine hours. During the time of this rise, the temperatures at the various stations throughout the spectrophotometer were noted. The results are shown in Fig. 7. The temperatures of the various parts of the source box followed very closely the room temperature which is shown by curve *A*. The temperatures of all parts inside the spectrophotometer, both metal and glass, excepting the prism, changed together and are shown by curve *B*. The prism temperature is shown by curve *C*.

These data show that during a period of five hours, the room temperature increased rapidly and the temperature of the parts inside the spectrophotometer changed very little. This lag in temperature rise has been effected by providing high resistance to heat flow and large thermal capacity. Since the temperatures of the parts inside the spectrophotometer change very slowly, the indication is that although the average room temperature should be controlled rather carefully, large fluctuations over a comparatively short time will have little or no effect on the accuracy of the instrument.

The second series of observations was made by turning the Nernst glower off for several days to allow the room and all parts of the spectrophotometer to approach as nearly as possible a temperature equilibrium. Then, with the slits open and set at 1.0 mm width and with the spectrophotometer set at 13.8 microns, the source was turned on, and temperature measurements were made periodically for seven hours. (Source operated at 65 watts.) At the end of two hours, it was found that temperature equilibrium had been established inside the spectrophotometer with the following increases in temperature:

Entrance slit jaws	3.7°C
Entrance slit support	1.2
First mirror	0.4
Prism	0.1
All other internal parts	0.0

Essentially the same results were obtained with 1.8 mm slits. These observations were made with no water cooling of the source box. It took seven hours for the source box and its parts to reach temperature equilibrium. The increases in temperature were as follows:

Source box walls	18.6°C
Shutter	25.6
Source mirror	13.6

The magnitude of the changes inside the spectrophotometer indicates that the energy entering the slit is of little importance in changing the temperatures of internal parts.

BAFFLES

The manner in which the temperatures of the interior parts are affected by energy entering the entrance slit depends to some extent upon the location of baffles, the arrangement of which is indicated by dotted lines in Fig. 2. Baffles have been used generously at all positions where they would help to prevent scattered radiation from

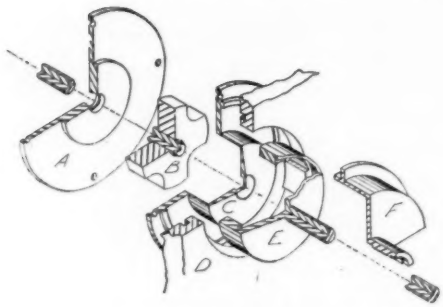

FIG. 9. Magnetic drive. *A*, bearing housing; *B*, four-pole Alnico No. 5 magnet; *C*, thin cylinder of No. 772 manganese alloy; *D*, section of vacuum tank; *E*, four-pole soft steel slave; *F*, bearing housing.

reaching the thermopile. Tests showed that one of the most important is the baffle which shields the jaws of the entrance and exit slits. This prevents radiation, which is returned by the Littrow mirror to the entrance slit jaws, from being reflected into the exit slit.

The baffles are made of aluminum sheet with a special non-reflecting surface. The aluminum was first sprayed with a slow-drying black lacquer and, while the lacquer was wet, "flocked" black wool was blown onto the surface. After the lacquer was dry the excess wool was brushed off. In order to minimize further reflection from the surface of the lacquer, a thin coat of bone black was then sprayed upon the "flocked" surface.

SCATTERED RADIATION

The amount of scattered radiation throughout the region from 9μ to 15μ is shown, in one way, by curve *A* in Fig. 8. This gives, for each wavelength, the voltage change in the thermopile output introduced when the metal shutter is replaced by the mica shutter. These readings were obtained with 1.8 millimeter slits for all positions.

Curve *B* shows what percentage of the total deflection at each position is due to scattered radiation. All points except the one at 4.3μ and the high value at 15μ were obtained by dividing the ordinate of curve *A* by the total deflection at the corresponding prism position, the total deflection being the change in the thermopile output which occurs when the metal shutter is opened (Slits at 1.8 millimeter). The point at 4.3μ and the higher point at 15μ were established somewhat differently. For these points the change in the thermopile output, which results when the metal shutter is replaced by a gas absorption cell filled with CO_2 gas, was divided by the change in thermopile output which occurs when the metal shutter is opened (no CO_2 in the absorption cell). When the CO_2 was in the optical path, the pressure was such that it gave 100 percent absorption at the two positions.

DRIVING MECHANISM

Two drives through the vacuum tank walls are necessary, one to rotate the prism and the other to adjust the slits. Greased cone seals are undesirable because of leakage and friction. Selsyn drives were ruled out because they would tend to develop heat inside the spectrophotometer. No hermetically sealed drives were commercially available, so a special magnetic drive was designed for this purpose.

The magnetic drive consists of a four-pole Alnico No. 5 magnet, *B* in Fig. 9, a four-pole soft steel slave, *E*, and a thin cylinder of No. 772 manganese alloy *C* (obtained from the W. M. Chace Company) in the gap between the magnet and slave. The thin cylinder provides a vacuum-tight enclosure. The Alnico magnet is driven by a synchronous motor on the outside of the vacuum tank and the steel slave is driven on the inside of the tank by magnetic attraction. The No. 772 alloy is a high-resistance non-magnetic material of high mechanical strength ideally suited to this application. A maximum torque of 120 ounce-inches, which is more than adequate for the intended operation, can be transmitted by this drive without slippage.

Beryllium-copper sylphons are used as universal couplings throughout the system. The sylphons are particularly applicable here for the following reasons:

1. They have small torsional flexibility.
2. As used in this instrument, they transmit constant angular velocity to the driven member.
3. They are small in size and of low inertia.
4. They have no rubbing parts to wear and develop back lash.
5. They permit mechanical driving of parts on the bed plate with very little end thrust.

The prism table drive was designed to have minimum wear and back lash. The gearing from

the magnetic drive to the prism table, which is a reduction of 129,600 to 1, is designed to have a minimum number of gears and to have them so placed that their inaccuracies will not affect the measurement of prism rotation. Most of the reduction is obtained in the fine pitch screw which drives a long arm fastened to the prism table as shown in Fig. 6. The nut on the screw is made of a tough bronze and is split for adjusting to minimum play. The screw is made of a high-carbon, high-chrome, stainless steel hardened to approximately 60 Rockwell *C*. The threads are ground with a precision thread grinder, accurate to 0.0001 in 4 inches with error accumulative. A weight and pulley system applies uniform pressure to the arm, nut, and screw thrust bearing to remove all back lash. Measurements made by means of a microscope show that back lash in the drive between the Veeder counters and prism is only about $\pm\frac{1}{2}$ Veeder number or $\pm\frac{1}{2}$ second of arc of prism rotation.

A tantalum-carbide sphere-and-plane thrust bearing is used at the end of the screw. Two tantalum-carbide cylinders at right angles are used as the contact between the nut and arm. This is an excellent bearing material and highly resistant to wear. The driving mechanisms have proven to be very satisfactory. The instrument has been in operation nearly a year, during which time about two hundred spectra have been recorded, and no change in calibration has been observed.

As has been done in many previous infra-red spectrophotometers, Veeder counters are used as a simple and effective method of indicating both prism position and slit width. One Veeder counter reads prism position in seconds of arc directly, except for small corrections near the ends of the screw. The other counter reads the slit width directly in microns. As a check on possible slippage of the magnetic drives, two Veeder counters are placed inside the spectrophotometer where they can be read through windows, and another pair on the outside of the magnetic drive in the control console. In six months of normal operation, no slipping has occurred.

The recording chart and the prism are both driven by synchronous motors. A solenoid-operated brake is automatically applied on the prism drive motor whenever the record is stopped in order to avoid coasting which would throw the two motions out of synchronism. The coasting at maximum recording speed does not exceed 2 seconds of arc of prism rotation.

As indicated in Fig. 10, the speed can be varied by a set of change gears. A sliding gear controlled by a knob on the panel, provides a rapid drive in either direction. Another control disengages the motor to allow manual adjustment of prism rotation from the panel.

SOURCE BOX

The source is a Nernst glower operating in a vacuum. The source box shown in Fig. 11 is designed as a cylinder to withstand atmospheric pressure and to simplify construction, particularly the machining of the vacuum seals at the front window and back plate. The cylinder is water jacketed for cooling. The source mirror is

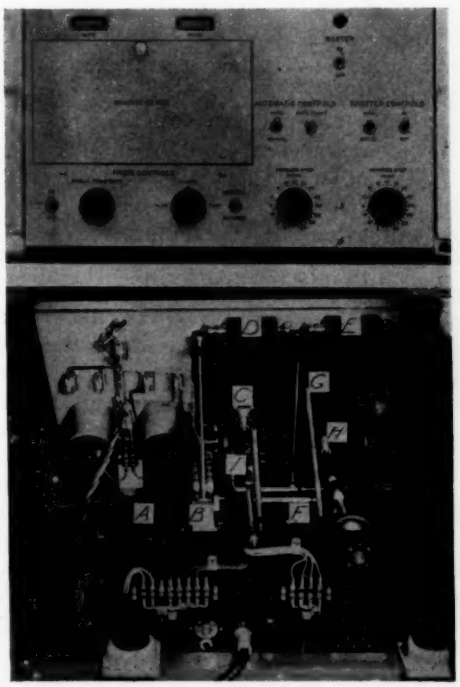

FIG. 10. Front and rear views of prism control panel. *A*, prism drive motor; *B*, automatic brake; *C*, solenoid for brake; *D*, veeder counter for prism; *E*, veeder counter for slits; *F*, gear train for quick return; *G*, change gears; *H*, manual drive for prism; *I*, reverse controls.

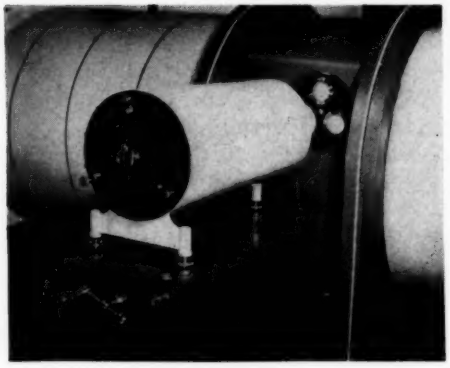

FIG. 11. Source box and absorption cells. The focus and position of the image of the source on the slit may be adjusted without breaking the vacuum.

mounted on the back plate. An image of the glower can be focused on the slit while the tank is evacuated. This is done by means of the three adjusting screws working through sylphons. A small glass window makes it possible to view the image of the source on the slit. The whole source box can be retracted and returned to a stop if desired. It is supported on three legs exactly like the arrangement described previously for the optical bed plate. Mica and metal shutters are located in the source box and are operated from the control console by means of solenoids. A portion of the source box close to the slit may be removed to allow the insertion of a 6″ long gas cell without changing the operating position of the source box. A thermopile in the source box is exposed to the Nernst glower; this permits periodic checks to be made on the total radiation intensity.

ABSORPTION CELLS AND TURRET

A three-position turret rotates in front of the window between the source box and the entrance slit of the spectrophotometer. Without disturbing the source box, the operator can insert an absorption cell into the turret and rotate the latter until the cell is located in front of the slit.

The absorption cell for liquids is designed to enable the operator to open or seal the ports for admitting samples without using any tools or removing any parts which might be dropped or lost. The ports are sealed by spring-loaded neoprene gaskets fastened to a door on the side of the cell which is held closed by a latch. Releasing the latch opens the door and the two ports as shown in Fig. 12. The cell is cylindrical, so that the salt disks and cell body can be readily and accurately machined in a lathe. The salt discs extend beyond the housing so that the outer surfaces can be relapped and polished without having to disassemble the cell. It might be pointed out that the same provisions are made for the windows in the source box and vacuum tank. The absorption cells are easily inserted or removed by means of a handle having a bayonette type lock. With this handle, fogging of the cell windows by handling is minimized. Figure 13 shows the gas cell which is designed to fit into the source box replacing a section of the source box normally used when thin liquid cells are used.

OPTICAL ADJUSTMENT

Proper adjustment of the optics of a spectrophotometer is a difficult operation unless a logical sequence is followed. No published description of the procedure for instruments of this general type has been found. It is known that various procedures have been followed, some of which are unnecessarily involved or unsound. A brief summary of the sequence used in the present instrument will therefore be presented.

It is assumed that the off-axis angle of the parabolic mirror is known and that the slits and all supports for optical parts are placed in the design position as accurately as is feasible with

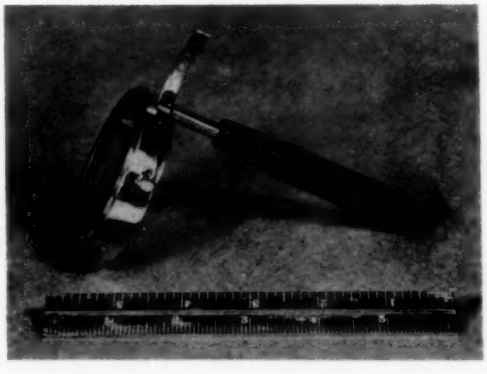

FIG. 12. Absorption cell for liquids. The filling ports are exposed by pulling back a latch opening a hinged cover.

good machine-shop practice. The axis of rotation of the prism is fixed and all alignment is made with respect to this axis.

1. Set the axis of a reference telescope (with Gausian eyepiece) perpendicular to the axis of rotation of the prism table. A method may be used involving successive approximate adjustments of level of a test prism temporarily mounted on the prism table and of the telescope trained successively on different prism faces.

2. Adjust the Littrow mirror until it is perpendicular to the axis of the telescope, that is, parallel to the axis of prism rotation. The accuracy of adjustments 1 and 2 influence all subsequent alignment and can readily be made to ± 5 seconds of arc.

3. Bring the axis of the paraboloid into the plane which is perpendicular to the Littrow mirror and which passes through the centers of exit slit and paraboloid. This involves focusing the exit slit back on itself. A crosswire may be placed across the slit, and the image of the crosswire, formed by the paraboloid and reflected from the Littrow mirror, may be observed with a microscope while three adjustments are made repetitively to successively better approximations:

A. Focal position (horizontal travel) of the paraboloid. This adjustment is made, at this point, only to facilitate part *C*.
B. Rotation around a horizontal axis parallel to the face of the off-axis section of the paraboloid.
C. Rotation around axis normal to the face. This adjustment can be made to $\pm 1°$ by careful microscopic examination of image definition.

4. Rotate paraboloid about vertical axis of mirror holder to bring axis of paraboloid to the design position shown in Fig. 2. An illuminated slit may be temporarily placed at *F* and all but a narrow, vertical slit of the paraboloid masked off. The Littrow mirror and paraboloid are rotated about vertical axes until the image of the test slit at *F* returns to *F*. The Littrow mirror is then perpendicular to the axis of the paraboloid and should be left there for steps 5 through 9. Steps 3 and 4 may now be repeated to improve adjustment further.

5. Rotate the entrance mirror to bring the image of the exit slit into coincidence with the entrance slit.

6. Adjust the focal position of the paraboloid to focus exit slit on entrance slit.

7. Rotate slits to parallelism.

8. Adjust the optical system in source box to line up with internal optical path.

9. Install a source of low-intensity (to avoid damage to thermopile) and adjust the exit mirror system to bring focused image of exit slit onto the target of the thermopile. Final adjustment of focus of the thermopile mirror may be made by observing maximum output from the thermopile.

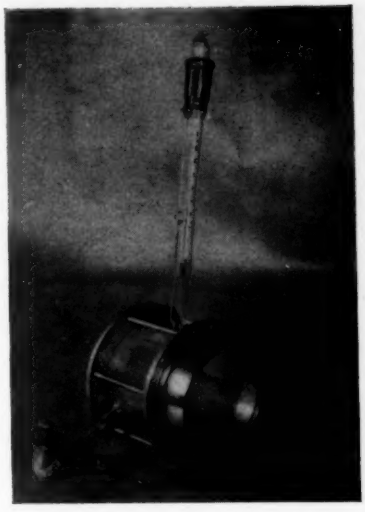

FIG. 13. Absorption cell for gases.

10. Install the salt prism on its table in calculated position with respect to Littrow mirror. Adjust prism around horizontal axis to bring faces perpendicular to reference telescope.

11. Using a sodium light source, rotate prism table to bring the *D* lines on the exit slit. Adjust prism driving system and Veeder counters to proper positions.

CONCLUSION

This spectrophotometer has been in operation for approximately one year during which time its convenience and speed have been demonstrated in routine analysis and research work. Recently some records of well known spectra of gases have been made to study, in more detail, such characteristics as resolution, sensitivity, and reproducibility. This study is discussed in another paper.

ACKNOWLEDGMENT

The authors are indebted to Mr. R. N. Frawley for contributions to the design and to Mr. R. E. Jutras for his assistance in the mechanical assembly of the instrument.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 11 NOVEMBER, 1947

The Refractive Indices of Alcohol, Water, and Their Mixtures at High Pressures

JOSEPH S. ROSEN[1]
University of Kansas City, Kansas City, Missouri
(Received June 30, 1947)

The effect of pressure up to 1800 kg/cm² on the refractive indices of ethyl alcohol, water, and five intermediate aqueous solutions of alcohol at 25°C, was measured for the Hg arc lines of wave-lengths 579, 546, 436, and 406 mμ. There is an approximate decrease of 0.6 to 1 percent in the Lorentz-Lorenz "constant," 0.1 to 0.45 percent deviation in the Gladstone and Dale "constant," and 0.2 to 0.6 percent deviation in the empirical Eykman formula. It is concluded that with none of these formulae can the compressions of the solutions be computed from the refractive-index measurements with any great accuracy.

The specific refractions of Gladstone and Dale are found most suitable for a mixture formula which holds as well through the pressure range of 1500 atmos. as at atmospheric pressure.

Compressions calculated from refraction formulae are found to be a linear function of the observed compressions of the solutions. An equation is given which reproduces with great precision the refractive indices of alcohol, water, and their mixtures; its logartihmic term is the same as that in the Tait equation for compressibilities, and the part containing the refractive indices involves the same function which appears in the formulae of Gladstone and Dale, Lorentz-Lorenz, etc.

THE effect of pressure on the refractive index of liquids had received considerable experimental attention before this century,[2] but no really high pressures were attained until recently.[3] The theoretical Lorentz-Lorenz relation between index of refraction and density,

$$(n^2-1)/(n^2+2)\rho = \text{constant},$$

has perhaps had its severest experimental trial through the comparison between values for the liquid and vapor states, where the density changed by a factor of as much as 1000. The agreement is surprisingly fair in spite of assumptions that cannot be expected to hold in both states. So, too, where the density is varied by pressure, there is to be expected a slight departure from the Lorentz-Lorenz relation.

The breakdown of the Clausius-Mossotti relation for polar substances led to the Debye-Langevin theory of dipoles and gave stimulus to much recent theoretical and experimental investigations on dielectric constants. Interest in the Lorentz-Lorenz relation, however, has lapsed except insofar as it is incidental to the calculation of electric moments. Similarly, interest in the mixture laws is primarily due to its relation to the binary solution method of measuring electric moments. The mixture laws of binary solutions give valuable information on the effects of molecules on the refraction of neighboring molecules. At atmospheric pressure the evidence is that intermolecular action upon refraction is negligible.[4] The work in the present paper on the mixture law under pressure extends these conclusions even to conditions where the molecules are brought closer and possibly distorted.

EXPERIMENTAL

The apparatus employed for measuring the index of refraction at high pressures has already been described;[5] it will only be briefly reviewed here. A heavy steel cylinder is fitted at its two ends with movable parts designed to apply and measure pressure. Prevention of leaks between these units is attained through packing, utilizing the "unsupported area" principle.[2]

In the pressure chamber between the upper and lower movable parts, two narrow slits permit the entry and emergence of light (see Fig. 1). These slits are continued through two movable steel lugs fitted against the inner wall of the pressure chamber. Two flat surfaces on the lugs were available for the mounting of windows.

[1] The experimental part of this paper is the joint work of F. E. Poindexter and the present writer. Only an abstract of the work has been published; this gives the constants of a function of the pressure representing the indices. See F. E. Poindexter and J. S. Rosen, Phys. Rev. **45**, 760 (1934).

[2] P. W. Bridgman, *The Physics of High Pressure* (The Macmillan Company, New York, 1931).

[3] P. W. Bridgman, Rev. Mod. Phys. **18**, 75 (1946).

[4] C. P. Smyth, *Dielectric Constant and Molecular Structure* (The Chemical Catalog Company, New York, 1931), p. 159.

[5] W. J. Lyons and F. E. Poindexter, J. Opt. Soc. Am. **26**, 146 (1936).

With these variable lugs, a maximum prism angle of about 51° was obtainable.

In an apparatus of this size and number of parts, it would have been extremely bothersome, if not impossible, to keep the compressed liquid free from contamination. Consequently, the liquid of which the index was measured was confined to a small rubber sack, with its two ends secured about the periphery of the thick, elliptic-shaped windows. The remainder of the cylinder was filled with glycerine which transmitted the pressure.

The refractive index was determined by the minimum deviation method using the conventional formula. Measurements were made with the lines of the mercury arc spectrum of wavelengths 579, 546, 436, and 406 mμ.

The refractive indices were measured at room temperature; the thermal effects of compression were effectively minimized by the massiveness of the apparatus and the long intervals between readings. Experiment showed that thermal equilibrium between liquid and apparatus was established in about 10 minutes for the pressure intervals taken in this investigation. It is not probable for this reason that temperatures varied by more than one degree at the time readings were taken. At 2000 atmos. this would affect the index of water by approximately .0001, and by .0002 for alcohol, with intermediate values for mixtures. The refractive indices shown in Tables I and II are the values obtained by adjusting the data to 25°C, by using experimentally determined values of dn/dt at various pressures. The refractive indices were measured[1] up to 1800 kg/cm² but are shown here, in 500-atmos. intervals, only to 1500 atmos., since compressibility data for alcohol mixtures are available only to this pressure.

COMPRESSIBILITY DATA

The specific volumes of water, ethyl alcohol, and their mixtures have been calculated from the compressibility data of Moesveld.[6] The compressions,[7] k_P, for the alcohol-water mixtures used in this experiment, and the compressibilities at atmospheric pressure are shown in Table III. The specific volumes at any pressure, v_P, were computed from the relation $v_P=v_0(1-k_P)$, where v_0 is the specific volume at atmospheric pressure.

The wide applicability of Tait's equation to represent compressibilities of liquids has been shown in recent years,[8] and it is used here to reproduce Moesveld's results.

Tait's equation

$$\frac{dk_P}{dP}=\frac{C}{B+P}, \tag{1}$$

as customarily used, involves two parameters C and B both independent of the pressure; B is a function of the temperature, concentration, and properties of the solution; while C is approximately independent of the temperature.

From Tait's Eq. (1), β_P, the true compressibility at any pressure P, is (the temperature remaining constant)

$$\beta_P=\frac{dk}{dP}=-\frac{1}{v_0}\frac{dv}{dP}=\frac{C}{B+P}; \tag{2}$$

so that, β_0, the true compressibility at atmospheric pressure ($P=0$), is C/B. In the integrated form this equation is

$$k_P=\beta_0 B\ln(1+P/B); \tag{3}$$

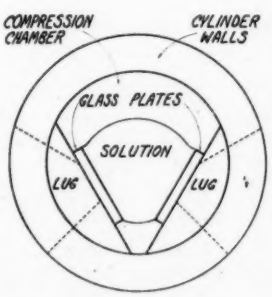

FIG. 1. Cross section of the apparatus used in measuring the refractive indices at high pressures.

[6] A. L. T. Moesveld, Zeits. f. physik. Chemie **105**, 450 (1923).

[7] The symbols used here are: k_P, the bulk compression, is $-(v_P-v_0)/v_0$; v_P is the specific volume at pressure P; $\beta_0=\left(-\frac{1}{v_0}\frac{dv}{dP}\right)_{P=0}$ is the compressibility at $P=0$ (atmospheric pressure); n_P is the refractive index at P atmospheres of pressure. The subscript 0 indicates atmospheric pressure. Concentration is expressed in percent by weight.

[8] P. W. Bridgman, Rev. Mod. Phys. **18**, 17 (1946). See also Harned and Owen, *The Physical Chemistry of Electrolytic Solutions* (Reinhold Publishing Corp., New York, 1943), p. 270.

where, in this modified Tait's equation, we have set $C=\beta_0 B$. For water, alcohol, and their mixtures Moesveld[6] gives for the compressibilities at atmospheric pressure

$$10^6\beta_0 = 44.5 - .5443c + .01754c^2 - .06585\cdot 10^{-3}c^3,$$

where c is the percent by weight of alcohol. Equation (3) is, in this instant, an especially convenient form as it requires the determination of only one parameter B, a not unimportant advantage when adjusting data to the equation by the method of least squares. Several other points about the relation $C=\beta_0 B$ are worth mentioning: since C is independent of the temperature (indeed, varies little for many liquids) the variation of B with temperature can be calculated if the temperature dependence of β_0 is known. Again, Eq. (3) readily shows how to effect extrapolations for β_0 from data at high pressures.

The results of adjusting the observed compressibility data of Moesveld for water, alcohol, and their mixtures to Eq. (3) by the method of least squares are shown in Table III.[9] Figure 2 shows the differences between the observed compressions of the alcohol, water, and their mixtures and those calculated by Tait's equation with the constants shown in Table III. Figure 3 shows the constant B of Tait's equation plotted against the concentration of the water-alcohol mixtures.

MIXTURE LAWS

The molar refraction, $R_{1,2}$, of a mixture with two components is defined by the equation[4]

$$R_{1,2} = \frac{(n^2-1)v}{n^2+2}(f_1M_1+f_2M_2),$$

TABLE I. The refractive indices and the specific refractions of alcohol, water, and their mixtures at different pressures at 25°C. $\lambda = 579$ mμ and 546 mμ.

Percent alcohol by weight	Pressure, atmos.	Specific volume	n_{5579}	Lorentz-Lorenz "constant" Obs'd.	Lorentz-Lorenz "constant" Calc'd.	Gladstone-Dale "constant" Obs'd.	Gladstone-Dale "constant" Calc'd.	n_{5461}	Lorentz-Lorenz "constant" Obs'd.	Lorentz-Lorenz "constant" Calc'd.	Gladstone-Dale "constant" Obs'd.	Gladstone-Dale "constant" Calc'd.
Water	1	1.0029	1.3330	.2063		.3340		1.3340	.2069		.3350	
	500	.9818	1.3401	.2059		.3340		1.3413	.2065		.3351	
	1000	.9633	1.3462	.2052		.3335		1.3474	.2059		.3347	
	1500	.9471	1.3516	.2046		.3330		1.3428	.2053		.3347	
19.80	1	1.0345	1.3464	.2205	.2211	.3584	.3586	1.3479	.2214	.2217	.3599	.3597
	500	1.0147	1.3534	.2203	.2207	.3586	.3589	1.3546	.2209	.2214	.3598	.3602
	1000	.9968	1.3592	.2196	.2200	.3581	.3586	1.3605	.2203	.2207	.3594	.3598
	1500	.9815	1.3645	.2190	.2193	.3578	.3580	1.3657	.2197	.2200	.3589	.3598
40.04	1	1.0737	1.3569	.2351	.2362	.3832	.3838	1.3583	.2360	.2369	.3847	.3849
	500	1.0505	1.3653	.2349	.2358	.3837	.3844	1.3665	.2356	.2365	.3850	.3857
	1000	1.0308	1.3720	.2343	.2350	.3835	.3843	1.3731	.2349	.2358	.3846	.3856
	1500	1.0139	1.3776	.2335	.2344	.3828	.3839	1.3787	.2341	.2351	.3840	.3855
60.80	1	1.1298	1.3625	.2509	.2517	.4096	.4096	1.3636	.2516	.2524	.4108	.4108
	500	1.0990	1.3738	.2509	.2513	.4108	.4104	1.3750	.2516	.2521	.4121	.4120
	1000	1.0748	1.3821	.2502	.2505	.4107	.4105	1.3832	.2508	.2513	.4140	.4120
	1500	1.0545	1.3889	.2493	.2498	.4101	.4102	1.3902	.2501	.2505	.4115	.4018
81.30	1	1.1963	1.3634	.2663	.2670	.4347	.4351	1.3645	.2670	.2678	.4361	.4364
	500	1.1547	1.3772	.2657	.2666	.4356	.4362	1.3786	.2666	.2675	.4372	.4379
	1000	1.1241	1.3874	.2649	.2658	.4355	.4365	1.3889	.2658	.2666	.4372	.4380
	1500	1.0994	1.3955	.2639	.2650	.4348	.4363	1.3971	.2648	.2658	.4366	.4378
90.74	1	1.2321	1.3628	.2738	.2741	.4470	.4468	1.3641	.2747	.2749	.4486	.4482
	500	1.1846	1.3779	.2730	.2736	.4477	.4480	1.3795	.2741	.2746	.4496	.4499
	1000	1.1505	1.3890	.2721	.2728	.4475	.4485	1.3901	.2729	.2737	.4488	.4500
	1500	1.1236	1.3980	.2712	.2720	.4472	.4482	1.3992	.2719	.2728	.4485	.4498
Alcohol	1	1.2738	1.3598	.2810		.4583		1.3609	.2818		.4597	
	500	1.2198	1.3769	.2805		.4597		1.3784	.2815		.4616	
	1000	1.1814	1.3895	.2797		.4602		1.3909	.2806		.4618	
	1500	1.1523	1.3992	.2789		.4600		1.4005	.2797		.4615	

[9] The assumption in Tait's equation that B is independent of the external pressure on the solution is justified by successful adaptation of this equation to compressibility data of many liquids (and some solids). Nevertheless, the values of B are not constant and depend on the pressure range of the data. Rightly, B is a function of the pressure.

TABLE II. The refractive indices and the specific refractions of alcohol, water, and their mixtures at different pressures at 25°C. $\lambda=436$ mμ and 406 mμ.

Percent alcohol by weight	Pressure, atmos.	Specific volume	n_{436}	Lorentz-Lorenz "constant" Obs'd.	Lorentz-Lorenz "constant" Calc'd.	Gladstone-Dale "constant" Obs'd.	Gladstone-Dale "constant" Calc'd.	n_{406}	Lorentz-Lorenz "constant" Obs'd.	Lorentz-Lorenz "constant" Calc'd.	Gladstone-Dale "constant" Obs'd.	Gladstone-Dale "constant" Calc'd.
Water	1	1.0029	1.3398	.2101		.3408		1.3422	.2115		.3432	
	500	.9818	1.3472	.2097		.3409		1.3496	.2110		.3432	
	1000	.9633	1.3531	.2089		.3401		1.3558	.2104		.3427	
	1500	.9471	1.3588	.2084		.3398		1.3613	.2097		.3422	
19.80	1	1.0345	1.3534	.2245	.2251	.3656	.3659	1.3559	.2260	.2266	.3682	.3685
	500	1.0147	1.3606	.2243	.2247	.3659	.3662	1.3631	.2257	.2261	.3684	.3688
	1000	.9968	1.3665	.2236	.2239	.3653	.3656	1.3691	.2250	.2255	.3679	.3685
	1500	.9815	1.3718	.2230	.2233	.3649	.3654	1.3744	.2244	.2247	.3675	.3680
40.04	1	1.0737	1.3643	.2395	.2405	.3912	.3915	1.3668	.2410	.2421	.3938	.3943
	500	1.0505	1.3727	.2392	.2401	.3915	.3921	1.3753	.2407	.2416	.3943	.3950
	1000	1.0308	1.3794	.2384	.2392	.3911	.3918	1.3820	.2399	.2409	.3938	.3948
	1500	1.0139	1.3852	.2377	.2386	.3906	.3916	1.3879	.2392	.2401	.3933	.3943
60.80	1	1.1298	1.3697	.2554	.2562	.4177	.4177	1.3720	.2568	.2579	.4203	.4208
	500	1.0990	1.3814	.2554	.2558	.4192	.4187	1.3840	.2570	.2575	.4220	.4218
	1000	1.0748	1.3897	.2546	.2549	.4189	.4186	1.3923	.2561	.2567	.4216	.4217
	1500	1.0545	1.3967	.2538	.2542	.4183	.4184	1.3994	.2553	.2558	.4212	.4214
81.30	1	1.1963	1.3705	.2709	.2718	.4432	.4437	1.3734	.2728	.2735	.4467	.4469
	500	1.1547	1.3844	.2702	.2713	.4439	.4450	1.3877	.2723	.2731	.4477	.4483
	1000	1.1241	1.3951	.2696	.2704	.4441	.4450	1.3982	.2714	.2723	.4476	.4484
	1500	1.0994	1.4034	.2685	.2696	.4435	.4449	1.4061	.2701	.2713	.4465	.4481
90.74	1	1.2321	1.3703	.2789	.2790	.4563	.4557	1.3727	.2805	.2807	.4592	.4590
	500	1.1846	1.3857	.2781	.2785	.4569	.4571	1.3888	.2800	.2803	.4606	.4605
	1000	1.1505	1.3965	.2768	.2776	.4562	.4572	1.3993	.2785	.2795	.4594	.4606
	1500	1.1236	1.4058	.2759	.2767	.4560	.4571	1.4087	.2776	.2785	.4592	.4604
Alcohol	1	1.2738	1.3669	.2860		.4674		1.3696	.2878		.4708	
	500	1.2198	1.3844	.2855		.4689		1.3873	.2874		.4724	
	1000	1.1814	1.3971	.2846		.4691		1.4001	.2865		.4727	
	1500	1.1523	1.4071	.2837		.4691		1.4100	.2855		.4724	

in which the f's and M's refer to the mole fractions and molecular weights, respectively, of the two components. If we replace $(n^2-1)/(n^2+2)$ by the expression $(n-1)$ introduced by Gladstone and Dale, then $R_{1,2}$ is sometimes said to define, instead, the molar refractivity. We shall, however, refer to $R_{1,2}$, containing either function of n, as the molar refraction.

Further,

$$R_{1,2}=f_1R_1+f_2R_2,$$

where R_1 and R_2 are the molar refractions of the pure components. It can readily be shown that this equation can be replaced by

$$r_{1,2}=c_1r_1+c_2r_2=c_1(r_1-r_2)+100r_2, \qquad (4)$$

where the c's refer to percent by weight of the components and the r's refer to the specific refractions of the mixture and its two components. For convenience in computations Eq. (4), rather than the equation involving molar refractions and mole fractions, will be used.

The specific refractions for mixtures, $r_{1,2}$, both those observed and those calculated by Eq. (4), are shown in Tables I and II, and the agreement between them is good. Thus, it may be said that the specific refraction of a mixture, at any given pressure, is a linear function of the percent of

TABLE III. The compressibilities, β_0, at atmospheric pressure and the compressions, k_P, for alcohol, water, and their mixtures at 25°C; the parameter, B, in the Tait equation computed from these data. Pressures are in atmospheres.

Percent alcohol by weight	$\beta_0\times10^6$	$k_{500}\times10^3$	$k_{1000}\times10^3$	$k_{1500}\times10^3$	B	$2.30259\beta_0B$
100.00	100.6	42.4	72.5	95.4	1125	.26060
90.74	90.3	38.5	66.2	88.0	1218	.25334
81.30	80.8	34.8	60.4	81.0	1331	.24762
60.80	61.5	27.3	48.7	66.6	1762	.24930
40.04	46.6	21.6	40.0	55.6	2745	.29453
19.80	40.1	19.2	36.4	51.2	4220	.38955
0.00	44.5	21.1	39.5	55.7	3606	.36949

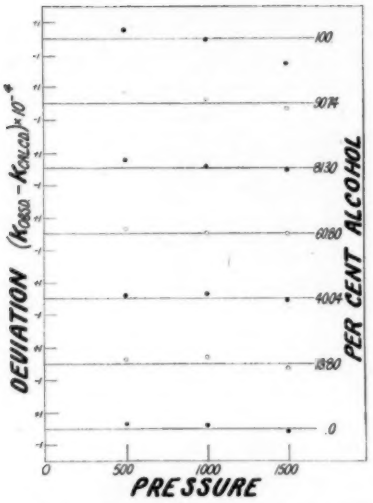

FIG. 2. The differences between the observed compressions of alcohol, water, and their mixtures and those calculated by Tait's equation with the constants shown in Table III.

alcohol (or water)[10] and, therefore, that the specific refractions of the pure components are constant.

Obviously, neither the Lorentz-Lorenz nor the Gladstone and Dale specific refractions remain constant with a change of pressure. In the data shown in Tables I and II there is an approximate deviation of .6 to 1 percent in the observed Lorentz-Lorenz "constant," and from .1 to .45 percent deviation in the Gladsone and Dale "constant." The empirical formula of Eykman, $(n^2-1)v/(n+.4)$="constant," which so successfully represents the refractive indices of the non-polar liquid benzene[11] to pressures of 1000 bars, was found to be slightly inferior to the Gladstone and Dale formula, but superior to the Lorentz-Lorenz formula. The Eykman "constant" (data not shown) shows deviations of .2 to .6 percent for alcohol, water, and their mixtures.

As a quantitative estimate of how closely the linear relation of Eq. (4) represents the specific refractions of the alcohol mixtures, we have proceeded as follows: we substituted into Eq. (4) $r_{1,2}$, r_1, and r_2, the *observed* values of the specific refractions (at the same pressure), and calculated the percent of alcohol for each mixture. In Fig. 4 the differences between the actual percent by weight of alcohol and those computed from the observed specific refractions are shown for various pressures. It is evident from this presentation that the linear relation is best approximated, for all pressures, by using the Gladstone and Dale specific refractions instead of the Lorentz-Lorenz refractions in the mixture law of Eq. (4). The same conclusion has previously been reached for these and other mixtures from data at atmospheric pressure.[12] The Eykman specific refractions in Eq. (4), as might be expected from what has previously been said, give calculated values for the concentration of the mixtures that are intermediate to those obtained by using the Gladstone and Dale and the Lorentz-Lorenz refractions.

REFRACTIVE INDEX FORMULAE

Of the important formulae for the refractive index of a solution which give a specific refractive "constant," all have the form $f(n)v=C$, where $f(n)$ is a function of the refractive index, v is the

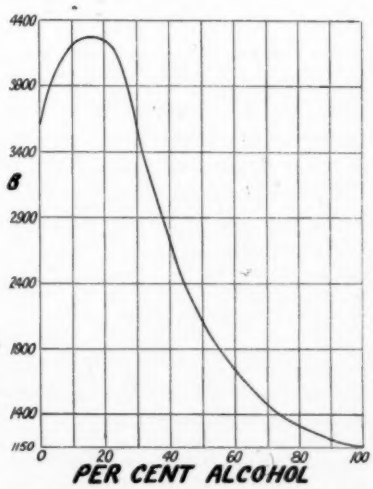

FIG. 3. The constant B of Tait's equation plotted against the concentration (percent by weight) of the water-alcohol mixtures at 25°C. The values of B are shown in Table III.

[10] The polarization of solutions of polar liquids such as alcohol and water is not expected to follow Eq. (4). It is, therefore, of interest to remark that Wyman has observed that the polarization of alcohol-water mixtures at 20°C is also very nearly a linear function of the mole fraction of either component. See J. Wyman, Jr., J. Am. Chem. Soc. **53**, 3301 (1931).

[11] R. E. Gibson and J. F. Kincaid, J. Am. Chem. Soc. **60**, 511 (1938).

[12] C. Dieterici, Ann. d. Physik **67**, 337 (1922).

specific volume and C the "constant." Where the formula is applied to solutions on which the pressure is varied, the expression

$$1-[f(n_0)/f(n_P)]=k_P' \qquad (5)$$

would equal the compression, k_P, if C remained constant. Thus the quantity k_P-k_P', the difference between the observed compression and that computed from Eq. (5), may be used as a measure in evaluating the worth of the several specific refraction formulae.

By this comparison the advantage generally is with the Gladstone and Dale refractive formula. At 1500 atmos. the compression of water calculated by Eq. (5) is about 5 percent too low by the Gladstone and Dale formula and about 14 percent below the observed value by the Lorentz-Lorenz formula. For alcohol, compressions calculated by either formula deviate as much as 7 percent, but the deviations are opposite in sign. The Eykman formula will reproduce the compressions of alcohol with discrepancies less than 2 percent, but for water the deviations are nearly 10 percent. It must be concluded that none of these formula are of much value for computing the compressions of the solutions of this experiment with any great accuracy.

A more important relation is disclosed on plotting k_P' of Eq. (5) against k_P; the points are collinear. Using the Tait Eq. (3) for k_P, we express this linear relation by the equation

$$1-[f(n_0)/f(n_P)]=m\log_{10}(1+P/B)+b, \qquad (6)$$

where m and b are constants and $f(n)$ is the Lorentz-Lorenz expression $(n^2-1)/(n^2+2)$, the Gladstone and Dale expression $(n-1)$, etc.

Some of the constants m and b of Eq. (6) were evaluated by the method of least squares and are given for alcohol, water, and their mixtures in Table IV. With these values in Eq. (6), the refractive indices of the solutions are reproduced with great precision for pressures above atmospheric. The values for n_P obtained from Eq. (6) for water ($\lambda=579$ mμ) deviate from the observed values given in Table I, at most, by one in the fourth place; the extreme deviation is 3×10^{-4} for alcohol.

TABLE IV. The constants m, m', and b of Eqs. (6) and (7) for $\lambda=579$ mμ. The parameter B is shown in Table III.

Percent alcohol	$f(n)=(n-1)$ m	b	$f(n)=(n^2-1)/(n^2+2)$ m	b
100.00	.25644	.00473	.22959	.00432
100.00	.28410*			
90.74	.24351	.00370		
81.30	.23580	.00412		
60.80	.23703	.00467		
40.04	.27303	.00335		
19.80	.35775	.00247		
0.00	.33846	.00191	.30622	.00179
0.00	.37018*			

* m' of Eq. (7).

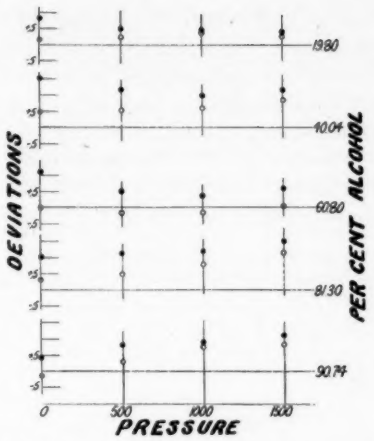

FIG. 4. Differences between the actual percent by weight of alcohol and those computed from the *observed* specific refractions of alcohol, water, and their mixtures. The circles indicate the deviations obtained by using the Gladstone and Dale refractions in the mixture law; shaded circles indicate deviations obtained by using the Lorentz-Lorenz specific refractions. The Eykman specific refractions give deviations which are approximately intermediate to those shown.

A second equation,

$$1-[f(n_0)/f(n_P)]=m'\log_{10}(1+P/B), \qquad (7)$$

may be used to represent the refractive indices of alcohol, water, and their mixtures at low pressures. This is the line $(k_P'=m''k_P)$ through the origin and through a point on Eq. (6), conveniently chosen at the lowest pressure for which the equation is still considered reliable. Thus, using the results of Eq. (6) for $P=500$ atmos., we obtain for alcohol and water ($\lambda=579$ mμ) the constants m' of Eq. (7) which are shown in Table IV.

Since direct verification of Eq. (7) for very low pressures is not possible with the data from the apparatus used in this experiment, the values of $(dn/dP)_{P=0}$ were computed from this equation

and compared with experimental results of previous investigators. For water ($\lambda = 579$ mμ, 25°C), Eq. (7) gives the value 14.8×10^{-6} for $(dn/dP)_{P=0}$, when $f(n)$ is both $(n-1)$ and $(n^2-1)/(n^2+2)$. This value agrees exactly with several independent observations.[13–15] This is also the value obtained from the Gladstone and Dale refraction formula with data at atmospheric pressure; and it confirms the observation of Quincke[15] that the Gladstone and Dale formula represents his experimental results for water at atmospheric pressure.

For alcohol ($\lambda = 579$ mμ, 25°C), Eq. (7) gives the value 39.5×10^{-6} for $(dn/dP)_{P=0}$ when $f(n)$ is $(n-1)$, and essentially the same value when $f(n) = (n^2-1)/(n^2+2)$. The available experimental data for alcohol are not in agreement;[15] the value extrapolated from Quincke[15] is about 37.6×10^{-6} and that of Röntgen and Zehnder[14] is 44×10^{-6}.

[13] *International Critical Tables*, Vol. VII, p. 13.
[14] W. C. Röntgen and L. Zehnder, Ann. d. Physik **44**, 24 (1891).
[15] G. Quincke, Ann. d. Physik **19**, 401 (1883); *ibid.* **44**, 774 (1891).

Book Review

Mathematical Analysis of Binocular Vision

RUDOLF K. LUNEBURG. Pp. 104+vi. Dartmouth Eye Institute by the Princeton University Press, Princeton, New York, 1947. Price $2.50.

This remarkable paper establishes the metric of the binocular visual space, i.e., the mathematical expression for the apparent size of a line element, and shows it to be non-Euclidean.

First, a unique psychometric coordination of numbers to spatial sensations is shown to exist. Following the introduction of special bipolar coordinates and a general discussion of a metric, the psychometric of visual sensations is established. The basic assumption is that an observer has a sensation of distance between two physical points and that this sensed distance depends on the coordinates of these points. For convenience, however, the symbolism of the infinitesimal calculus is used. By ingeneous methods and observations the general expression for the metric is reduced to a form which corresponds to one of three geometries, the elliptic, the Euclidean, or the hyperbolic. The final choice is made on the basis of the demonstration that the hyperbolic geometry alone allows the observer to improve his size judgment by training.

An independent second derivation of the metric is given by analyzing an interesting observation in the horizontal plane. In further support of the hyperbolic metric it is shown that physical observation, as distinguished from visual perception of our environment, also leads to a hyperbolic metric, if discussed from a relativistic point of view.

Finally, it is demonstrated that the theory explains two well-known phenomena of binocular vision, namely, the frontal-plane horopter and the alley problem, and that it provides a satisfactory mathematical analysis of certain important discoveries of A. Ames.

PAUL BOEDER
American Optical Company

Letters to the Editor

Cathodoluminescence Emission Spectra of Zinc-Oxide Phosphors

R. E. SHRADER AND H. W. LEVERENZ
Radio Corporation of America, RCA Laboratories Division, Princeton, New Jersey
August 22, 1947

IT has been reported[1-3] that pure, dry-process zinc oxide (prepared by burning zinc in oxygen, and designated by hex.-ZnO) emits a narrow band, peaked near 3850A, whereas pure zinc oxide which has been heated to about 1000°C in a reducing atmosphere (e.g., heated in CO or H_2, and designated by hex.-ZnO:[Zn]) emits a broad band, peaked near 5050A, when excited by cathode rays (C.R.). During experiments with a projection C.R. tube screened with hex.-ZnO:[Zn], O. H. Schade found appreciable emission of ultraviolet (u.v.), *but only when the* C.R. *beam was well focused* (high current density). Schade reported, also, that there was no detectable persistence of the u.v. component after about 10^{-7} second, whereas the decay of the visible component is detectable for over 10^{-6} sec. (ϵ^{-1} value $\approx 10^{-6}$ sec.). In later experiments conducted by the writers in conjunction with F. H. Nicoll and A. Rose, the u.v. emission from hex.-ZnO:[Zn], under 30-kv C.R., was found to have the same spectral distribution as that reported for hex.-ZnO.

The upper part of Fig. 1 shows the u.v. and visible emissions of zinc-oxide phosphors. Also shown is the transmission curve of a Wratten No. 18A filter, which was used with a 931A multiplier photo-tube to separate most of the u.v. emission from the visible emission. The visible emission was measured with a 931A+Wratten No. 60 filter, or with an eye-corrected Weston photronic cell. It has been determined that hex.-ZnO phosphors (formed by burning zinc in excess oxygen) emit only the u.v. band under C.R. excitation, whereas hex.-ZnO:[Zn] phosphors may be made so they emit either (1) little or no u.v. at any current density, or (2) more intense u.v. than visible radiation at high current density, although in both cases there is efficient visible emission at all current densities.

Figure 2 shows that the u.v. band of a particular hex.-ZnO:[Zn] phosphor (No. A-1-11-11) predominates at current densities higher than about 0.1 amp. cm^{-2} at 30 kv (aluminized screen[3]). This was verified by measuring the spectral distribution from 4000 to 7000A, at low and high current densities, and finding that the ratio of rate of emission at 4100A to that at 5050A could be varied from less than 1/20 to over 2 by varying the current density (4100A is the shortest wave-length at which accurate measurements may be made with our recording spectro-radiometer[4]). By making measurements of luminescence output during a period when the phosphor screen was allowed to warm from about 50°C to 250°C, by heat dissipation from a well-focused scanning 30-kv 735-μA C.R. beam, it was found that the efficiency of the u.v. emission decreased faster than that of the visible emission as the temperature of the luminescing phosphor increased (Fig. 3). The efficiency decreases were not permanent, and so serve as an indication of the relative effect of temperature on the two luminescence processes.

One of the writers (R.E.S.) reasoned that the dependence of u.v. emission on current density might be explained by the presence of an absorption band which coincided with the u.v. emission. If there were *internal* u.v. emission at all current densities and a non-radiative absorption of sufficient oscillator strength, then the *external* u.v. emission would be observable only at high current densities when the emission density saturates the non-radiative absorption process. The probability of absorption of u.v. photons is increased in that the average optical pathlength of photons trying to escape from a small crystal of high refractive index (hex.-ZnO index = 2) may be longer than the diameter of the crystal. In the actual case of a C.R. tube, where the screen thickness is set for optimum efficiency of the visible luminescence, the absorption of u.v. is further increased on transmission through the fine-crystal phosphor coating from the C.R.-input side of the screen (the point of highest current density). Experimental verification of

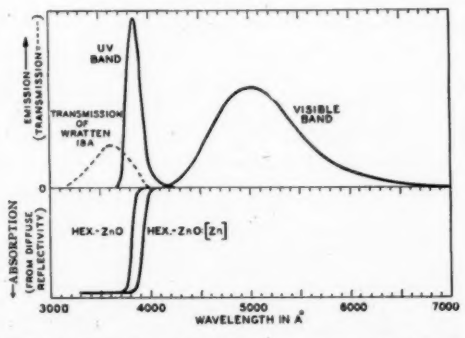

FIG. 1. Absorption and emission spectra of zinc-oxide phosphors.

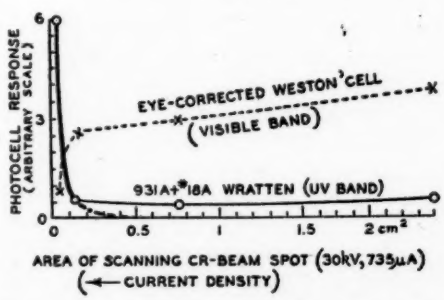

FIG. 2. Relative outputs of the two emission bands of hex.-ZnO:[Zn] as a function of C.R. current density (spot-size data from F. H. Nicoll).

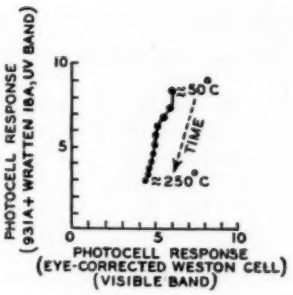

FIG. 3. Variation of cathodoluminescence efficiency as hex.-ZnO:[Zn] screen warmed during excitation by well-focused scanning C.R beam. 30 kv, 735 μA, scanned raster = 3 in. × 3 in.

the above postulate was obtained by measuring the diffuse reflectivity of hex.-ZnO and hex.-ZnO:[Zn] with the results shown in the lower part of Fig. 1. This postulate can account also for the apparent rapid decay of the u.v. emission, without requiring an extremely short lifetime of the excited state responsible for the u.v. emission. Just as the u.v. emission appears only when the *internal* u.v. emission density saturates the u.v. absorption, so the externally evident u.v. emission should terminate when the internal emission density during phosphorescence decreases to a value below that required to saturate the u.v. absorption process. Experimental proof of this point would be provided by decay measurements, of high time resolution, showing a downward break in the decay curve (L *vs.* t). Verification may be obtained by determining that the persistence of the u.v. band from hex.-ZnO (no visible band) is detectable over a longer time than that from hex.-ZnO:[Zn] (visible and u.v. bands). Further points of interest are: the visible emission of hex.ZnO:[Zn] is excited by the 4047A line of mercury, and the u.v. emission from one zinc-oxide phosphor readily excited visible emission from another adjacent hex.-ZnO:[Zn] phosphor.

[1] H. W. Leverenz, "Phosphors *versus* the periodic system of the elements," Proc. I.R.E. **32**, 256 (1944).
[2] H. W. Leverenz, *Final Report on Radar Indicators*, PB 25481 (Office of the Publication Board, Washington, D. C., 1945).
[3] H. W. Leverenz, "Luminescence and tenebrescence as applied in radar," RCA Review **7**, 199 (June, 1946).
[4] V. K. Zworykin, "An automatic recording spectroradiometer for cathodoluminescent materials," J. Opt. Soc. Am. **39**, 84 (1939).